General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

X-622-68-399
PREPRINT

STATISTICAL SPECIFICATION OF THE 500-MB HEIGHT FIELDS USING SMOOTHED MEDIUM-RESOLUTION RADIOMETRIC FIELDS OF NIMBUS II

FRANK L. MARTIN

OCTOBER 1968

GSFC

GODDARD SPACE FLIGHT CENTER

GREENBELT, MARYLAND

X-622-68-399
Preprint

STATISTICAL SPECIFICATION OF THE 500-MB HEIGHT FIELDS USING SMOOTHED MEDIUM-RESOLUTION RADIOMETRIC FIELDS OF NIMBUS II

Frank L. Martin[1,2]

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland 20771

---

[1]On temporary leave from the Naval Postgraduate School, Monterey, California, as a National Research Council - National Academy of Sciences Senior Postdoctoral Resident Research Associate, assigned to Goddard Space Flight Center.

[2]Support by the Navy Weather Research Facility, NAVAIR SYSTEMS COMMAND, is gratefully acknowledged for this work, most of which was completed during 1967-68, while the author was at his permanent post at the Naval Postgraduate School.

CONTENTS

STATISTICAL SPECIFICATION OF THE 500-MB HEIGHT FIELDS USING SMOOTHED MEDIUM-RESOLUTION RADIOMETRIC FIELDS OF NIMBUS II

by

Frank L. Martin

ABSTRACT

Over an area spanning the North American continent from tropical to polar latitudes, and between 40W and 135W, four consecutive days of the NIMBUS II MRIR five-channel equivalent blackbody temperatures, and of 1000 mb geopotential fields were used to specify the 500 mb field of geopotential. Three of the four days were reserved for dependent data, and the fourth day was reserved for independent test-case data. The data samples were stratified into "extratropical" and subtropical-tropical" areas, and then pooled three-day stepwise, screening regression equations were developed by areas. In the lower latitude stratifications, the 500 mb regression equations were only of marginal significance, but in the extratropical zone (latitudes 32-64 N), the regressions gave 500 mb height specifications with multiple correlation coefficients of 0.95. Moreover, in the independent-data test case, no shrinkage of explained variances occurred in these extratropical latitudes. Here the most important predictors for Z_5 (I, J) were the equivalent blackbody temperatures (T_{BB}) in the 14-16 micron and 10-11 micron channels, the 1000-mb geopotential, and finally an effective T_{BB} associated with the solar-reflectance channel.

1. INTRODUCTION

In an earlier unpublished study,[3] Martin and Warnecke (1967) investigated the feasibility of making a statistical determination of the 500-mb field of geopotential using nearly synoptic, composite medium-resolution radiometric grid print maps resulting from observations made during subtracks 0636, 0637, 0638, 0639 and 1640 of TIROS IV. These grid-print maps, provided by the National Aeronautics and Space Administration from their TIROS IV Final Meteorological Radiation (FMR) tapes, displayed the composite fields of the 6.3 micron water-vapor channel, the 8-12 micron window channel, and the 0.2-6.0 micron solar-reflectance channel of the TIROS system (Staff Members, 1963). All radiometric grid-print charts employed had a scale of 1: 10,000,000. The geographic area of the composited grid-point values in the TIROS experiment was essentially from the eastern part of North America to 140W longitude. Of the three radiometric channels available in composite grid-print form, the window channel gave the primary component of the explained variance of Z_5 (I, J), (the 500-mb height field), for the single test case examined, 00GMT, 25 March 1962, which was the map time most nearly time-centered relative to the suborbital tracks under consideration.

A somewhat different method was used by Jensen et al. (1966) to specify the 500-mb field by TIROS IV window-channel field observations. For each of a set of 26 contour grid-points, they sought statistical specifications Z_5 (I , J) in terms of a spatial distribution of nearly equally-spaced window-channel grid-point data. The statistical screening technique which they used is based upon that adapted by Miller (1962) to meteorological problems. While Jensen et al. used a

[3]Entitled "Some statistically-derived relationships involving medium-resolution radiation and conventionally-analyzed data fields."

total of 45 synoptic maps of radiation and contour fields they made use of only 26 contour grid-points spaced at five degree latitude – longitude intervals to specify the height field Z_5 (I , J). The results obtained by Jensen et al. (1966) gave approximately the same degree of specification of Z_5 (I , J) as indicated in the one-day preliminary study by Martin and Warnecke (footnote 3). In the latter study Z_5(I , J) was sought as a statistical linear function of the three equivalent blackbody temperatures (one for each channel) specified for the same (I , J). Values of (I , J) in the Martin-Warnecke study spanned the entire grid, covering approximately the same geographic area as that employed by Jensen et al., but in grid intervals of one degree latitude by longitude.

The statistical study just cited of the TIROS IV day 24-25 March 1962, led to the present more extensive project of expressing Z_5 (I, J) statistically in terms of the composite values of the five-channel MRIR NIMBUS II data at identical grid-point (I, J) values. All radiometric grid-point charts employed a Mercator projection upon which was superimposed a square-mesh grid, having a mesh interval of 1.25 degrees of longitude (cf. Fig. 1). The geographic coverage of the (I, J) grid-map for all I and J used here is shown in Fig. 1. A description of the radiometric channel-readings is discussed further in Section 3 of this study and considerable additional detail is given in the NIMBUS II User's Guide (Staff Members, 1966b, Sec. 4). Because of the greater selection of channels available in this study, in addition to the superior quality of the NIMBUS II data, a more satisfactory experiment has been anticipated than in the earlier one using TIROS IV radiometric data.

Several other improvements over the earlier Martin and Warnecke study have been incorporated in the present study. One such improvement is that the use of the regression equations have been stratified by geographic areas, in

particular those depicted in Fig. 9 of this paper. In this regard the best statistical results were found for the latitude zone 32N to 64N, with relatively poor specification on the equatorward side of this line of separation. In addition, the present view of the problem has been enlarged to one of specification of the thickness $h(I, J) = Z_5 - Z_{10}$ (Z_{10} = 1000-mb contour height) in terms of the column values of equivalent blackbody temperatures T_1, T_2, T_3, T_4, T_5, where the subscript identifies the channel - readings with which these temperatures are associated (Staff Members, 1966b), and T_5 is to be defined in Section 3.

As a result the independent variables for the stepwise regression of $Z_5(I, J)$ have been taken as (Z_{10}, T_1, T_2, T_3, T_4, T_5) with the regression coefficients selected by the Miller stepwise screening procedure (1962). The difference of the coefficient of Z_{10} from unity is viewed as an "air-mass conditioner" for the simultaneous existence of the radiometric properties T_1, T_2, T_3, T_4, T_5. As noted earlier, these radiometric temperatures are described in greater detail in Section 3. The data period involved in this experiment is listed in the following section.

2. THE DATA SOURCES FOR THE NIMBUS II MRIR DAYTIME PERIOD, 15-18 JULY 1966

Three successive days of dependent or test-data have been selected, while a fourth consecutive day has been reserved for independent sample data. The radiative parameters for the four listed MRIR days were coded onto IBM cards from initial composite grid-print maps provided by the National Aeronautics and Space Administration (NASA).

TABLE 1. Data-field identifications by days and their sources

Day number	1000-, 500-mb contour-map day	Source of radiometric data	
		MRIR day	Orbits averaged
1	00 GMT, 16 July 1966	15 July 1966	0816-0819
2	00 GMT, 17 July 1966	16 July 1966	0829-0832
3	00 GMT, 18 July 1966	17 July 1966	0842-0845
4	00 GMT, 19 July 1966	18 July 1966	0856-0859

Each numbered day in Table 1 corresponds to an "MRIR day" (cf., Staff Members, 1966a), within which the indicated numbered orbits occurred. The outgoing filtered radiances sensed by channels 1, 2, 3, 4, and 5 of the MRIR during these orbits constituted the radiative data set to be analyzed for that particular day. Contour-day numbers are identified by each of the four consecutive Fleet Numerical Weather Facility data fields for the indicated date at 00GMT.

With reference to Table 1, the data for each MRIR day were presented in the form of composite values from each of five channels during the orbits indicated, on a NASA grid-print chart (Staff Members, 1966b) containing grid points corresponding to each (I, J) of the Mercator map shown in Fig. 1. The data involved for the paired MRIR and contour days extend from 5N latitude (J = 01) to 64.3N latitude (J = 65), and from 135W longitude (I = 01) to 40W longitude (I = 77). A channel composite-value at (I, J) represents an average of all space-uncontaminated scan spots viewed by any of the five radiometers within an area of a grid square centered on (I,J). It should be noted that each MRIR composite orbital day corresponds to near-local noon solar times along the southern boundary of Fig. 1, and thus to approximately 1400 GMT where I = 77, varying progressively to about 2000 GMT at I = 01. These progressive time-differences relative to 00GMT should be borne in mind when the regressions relating Z_5 and Z_{10} with the five MRIR composite variables at the same (I, J) are carried out. With this understanding of the real-time differences, composited radiometric data will be identified by an MRIR day, and the four consecutive MRIR days used here are listed in Table 1 (after Staff Members, 1966a).

The IBM code for card-punch listing of radiometric and contour values covered each specific field in the form indicated below for each row-array of 77 entries, e.g. for J = 01:

J01	I01	followed by 18 four-digit temperatures (or contours)
J01	I19	similar card format
J01	I37	similar card format
J01	I55	similar format
J01	I73	followed by 5 four-digit temperatures (or contours).

A similar coding was employed for all other J -values progressively scanning the field northward. The radiometric values were actually rounded to three digits in equivalent degrees Kelvin, but prefixed by an initial zero as a spacer (e.g., 273K was coded as 0273).

Values of contour height Z_5 and Z_{10} were interpolated from hemispheric data tapes provided by the Fleet Numerical Weather Facility (FNWF) for the four contour-height days. FNWF contours are routinely listed on magnetic tapes at their respective NMC grid-point values (i , j), from which latitude φ and longitude λ are derivable from the polar stereographic map relationships

$$\sin \varphi = \frac{973.71 - [(i-32)^2 + (j-32)^2]}{973.71 + [(i-32)^2 + (j-32)^2]} \qquad (1)$$

$$\tan (\lambda + 10^\circ) = \frac{-(j-32)}{(i-32)} \qquad (2)$$

$$i, j = 1, 2 \ldots\ldots\ldots\ldots\ldots, 63$$

Here λ is considered positive in degrees west of Greenwich in the FNWF coordinate system.

The grid system (I , J) of Fig. 1 is expressible in terms of (λ , φ) of the Mercator-mapping (true at 22.5N longitude) by means of the relationships

$$\lambda = 135° - (I - 1)\ (1.25°)$$

$$\sin \varphi = \frac{2.23985\,[\exp .043634(J - J_0)] - 1}{2.23985\,[\exp .043634(J - J_0)] + 1} \qquad (3)$$

where $J_0 = 15.4753$ is the value of J corresponding to $\varphi_0 = 22.5$N, [for proof of (3) see Appendix]. A subroutine listing values of (I, J) in terms of (λ, φ) was developed and the results stored for future interpolative use. The duality of the (I , J) and (λ , φ) scales is indicated in Fig. 1 where the longitude-latitude scales are shown along the top and right-hand map borders, respectively.

By combination of (3) with (1) and (2), the (I , J) matrix equivalent of the FNWF coordinate grid mesh (i , j) was computed and stored. Since the fields Z_5 (i , j) and Z_{10} (i , j) were known, it was only necessary to employ the FNWF double-Bessel interpolation subroutine to determine the corresponding Z_5 (I , J), Z_{10} (I , J) fields. These interpolated Z_5 and Z_{10} fields, after being rounded off to the nearest whole meter, were printed on the Mercator maps of Fig. 1. The contour fields were comparatively smooth compared to the radiometric fields since the former fields had been previously analyzed using the relatively course NMC grid-mesh interval d = 381 km (true at 60N). On the other hand, the grid mesh of the NASA Mercator radiometric maps was of size D = 128.4 km at 22.5N, but represents an actual earth-size D cos φ km at latitude φ. (Here D is identical to ΔX given in Eq. A-2 of the Appendix).

In addition both Z_5 and Z_{10} (I, J) were punched on IBM cards to the nearest whole meter as four-digit field entries in five-card sets for each J-row, as described just before Eq. 1. The 3-digit 1000 mb heights were preceded by a zero-filler when the height was positive; however when the 1000 mb height was negative, the sign replaced the zero fill-digit. It should be observed now that within the data region of Fig. 1, there were no data voids in either of the height fields. However, due to certain boundary effects in connection with some of the orbits to be used for scanning there were grid points (I, J), the same for each channel, where no radiometric data were recovered. Any (I, J) with a radiometric data-void was card punched as 0000. For each day, all the original data values $\tilde{Z}_5$, $\tilde{Z}_{10}$, $\tilde{T}_1$, $\tilde{T}_2$, $\tilde{T}_3$, $\tilde{T}_4$, $\tilde{T}_5$ were coded so that the available entry at (I, J) occurred in identical sequences for each field. Note that the wavy superiod bar denotes original-field data, as contrasted with smoothed-field data to be described in Section 4.

3. THE FORM OF THE RADIOMETRIC FIELDS

The details of the infrared scanning in each of channels 1, 2, 3 and 4, the data-storage, and subsequent tape readout upon command from the data acquisition facility at Gilmore Creek, Alaska, are all clearly described in the NIMBUS II User's Guide (Staff Members, 1966b, Sec. 4). The latter reference also describes calibration procedures for conversion of the filtered infrared radiances into equivalent blackbody temperatures. The satellite orbital and scanning geometry, as listed on the Nimbus Meteorological Radiation Tape (NMRT) for each orbit then makes possible the assignment of a unique latitude and longitude to the point of earth scanned. The additional requirement of a radiometric nadir angle less than 50 degrees ensures that none of the radiometer data is space-contaminated.

For a Mercator grid print chart having a scale 1: 10,000,000, the population of scan spots per unit grid square varies in general from 5 to 15 over the major portion of the chart, smaller densities applying in general to the northern portions of the field where the grid mesh corresponds to a smaller earth area. Also, fewer scan spots occur when the sensing is accomplished by a single orbital swath, rather than by more or less overlapping swaths, viewed from two successive orbits. The NASA composite grid-print chart essentially lists the grid-square mean equivalent blackbody temperature based upon the recorded population count, for each of channels 1, 2, 3, 4. The composite value is then ascribed to the grid-point (I , J) at the center of the grid-square area. These temperatures will be denoted $\tilde{T}_1$, $\tilde{T}_2$, $\tilde{T}_3$, $\tilde{T}_4$. The spectral definition of the channel wavelimits is described in Section 4 of the Nimbus II User's Guide (Staff Members, 1966b), and these limits are briefly reviewed here:

Channel 1; 6.4 to 6.9 microns
Channel 2; 10 to 11 microns
Channel 3; 14 to 16 microns
Channel 4; 5 to 30 microns

For channel 5 (encompassing 0.2 to 4.0 microns), NASA provided for each day and grid point scanned, a composite reflectance also averaged in the manner just described to give the mean reflectance at (I,J). The Lambertian solar reflectance $R'(I', J', \theta')$ from a scan spot (I', J'), where the solar zenith angle is θ', may be defined as follows

$$R' = \pi N(I', J', \theta')/(S \cos \theta') \qquad (4)$$

In (4), N (I′, J′, θ') is the calibrated reflected solar intensity[4] in watts m^{-2} $ster^{-1}$ sensed by the solar-band scanning radiometer. The denominator of (4) is the undepleted incident solar intensity when the solar zenith angle is θ'. Values of θ' are known for each scan spot from the NMR taped listings.

The composite reflectance at grid point (I , J) of the NASA channel 5 grid-print map then averages all R′ -values, and in effect normalizes the zenith angle dependence in accordance with the formula

$$\tilde{R}(I, J) = \frac{\pi}{\eta S} \sum_{\eta'=1}^{\eta} \left[\frac{N(I', J', \theta')}{\cos \theta'} \right] \qquad (5)$$

where η is the population density of all scan spots (I′, J′) lying within a grid-square centered on (I , J). The term within the brackets on the right side of (5) has been called the adjusted intensity by Ruff et al. (1968). With the composite normalized reflectance $\tilde{R}(I, J)$ given by (5) as the output of the NASA channel

[4]The radiance N (I′, J′, θ) sensed by channel 5 has been multiplied by the factor 1.6513 in order to account for calibration by a solar constant S = 1395.0 watts m^{-2}, rather than its filtered value S* = 844.8 watts m^{-2} (Staff Members, 1966b, p. 56).

5 grid-print chart, one may solve for a normalized equivalent blackbody channel 5 composite temperature $\tilde{T}_5\ (I, J)$ in accordance with the Stefan-Boltzmann law (expressed in terms of reflected intensities):

$$\frac{\sigma}{\pi}\,[\tilde{T}_5\,(I, J)]^4 = \frac{S}{\pi}\,\tilde{R}\,(I, J) \tag{6}$$

Alternatively $\tilde{T}_5$ may be computed from (6) by means of

$$\tilde{T}_5 = \left(\tilde{R}\,\frac{S}{\sigma}\right)^{1/4}$$

In the last equation, $\tilde{R}$ is given in three-digit format (with a factor 10^{-3} suppressed) as printed on the NASA channel 5 composite reflectance chart. With the Stefan-Boltzmann constant $\sigma = 5.6687 \times 10^{-8}$ watts/m^2/(deg. K)^4, and S as listed in footnote 4, we arrive at the desired channel 5 equivalent blackbody temperature, $\tilde{T}_5$,

$$\tilde{T}_5 = 70.432\ [\tilde{R}(I, J)]^{1/4} \tag{7}$$

rounded off to a 3 decimal places.

From a hypsometric viewpoint, there is an advantage in having all radiometric variables expressed as equivalent blackbody temperatures. From a strictly

pragmatic point of view, the multiplicative range of $\tilde{T}_5$ as defined by (7) is only a factor of 2.4 (that is, from about 110K to 350K), whereas that of $\tilde{R}$ is at least a factor of 30. Hence map-contouring of $\tilde{T}_5$ is not much more of a problem than involved with the other four MRIR channels discussed earlier.

A problem in the regression analysis (Section 5) will be encountered, when it is found that in the (I, J) domain of Fig. 1, data gaps characteristically occur in the northwest corner of the radiometric fields whereas values of the contour heights $\tilde{Z}_5, \tilde{Z}_{10}$, are available everywhere during the data period. The data gaps in the MRIR data fields occur at identical locations on any one of the MRIR days but vary somewhat from day to day.

Because of the reduced grid size of the MRIR fields, their map features include some mesoscale detail, whereas the contour fields are of synoptic scale. As a result, it has been found convenient to perform a smoothing process, similar to that described by Schuman (1957), on all fields included in the regression analysis. This process is similar also to that performed on the regression variables in the Martin-Warnecke study [(1967); cf., footnote 3].

4. THE SMOOTHING PROCESS

The first description given here will be particularly applicable to the contour height fields over the full rectangular grid of Fig. 1, which has equally-spaced grid intervals of size D in both directions. For such fields, it is possible, in general, to define a Fourier representation with respect to wave number in both the $\underline{x}$-and $\underline{y}$-directions relative to the grid. Such a representation is of the form

$$\tilde{Z}(x, y) = \frac{1}{4} \sum_{|N| = 0}^{76} \sum_{|M| = 0}^{64} A_N \exp\left(\frac{2\pi i x}{ND}\right) B_M \exp\left(\frac{2\pi i y}{MD}\right) \qquad (8)$$

Here the harmonic function in the $\underline{x}$-direction is indicated as having wave number $2\pi/ND$, while the typical $\underline{y}$-feature is also harmonic and of wave number $2\pi/MD$, where

$$N = 0, \pm 2, \pm 3, \pm 4, \cdots\cdots\cdots\cdots, \pm 76$$

$$M = 0, \pm 2, \pm 3, \pm 4, \cdots\cdots\cdots\cdots, \pm 64$$

Corresponding wave-components have amplitudes A_N, B_M, respectively.

The simplest smoothing process described by Schuman (1957) involves a combination of weighting the $\underline{x}$-features at the points $\underline{x}$-D, $\underline{x}$, $\underline{x}$+D, using relative weight factors of 1 - 2 - 1, followed by a similar $\underline{y}$ -smoothing using identical weight factors on successive $\underline{y}$-points. This assumes that each point ($\underline{x}$, $\underline{y}$) is not on a boundary parallel to the $\underline{y}$ -axis when only $\underline{x}$-smoothing is performed. The result of a single $\underline{x}$-smoothing has the effect of replacing $\tilde{Z}(\underline{x}, \underline{y})$ by

$$^1\tilde{Z}^x = \frac{1}{4} \sum_{|N|} \sum_{|M|} A_N \exp\left(\frac{2\pi i x}{ND}\right) \left\{\left[\exp\left(\frac{2\pi i}{N}\right) + 2 + \exp\left(\frac{-2\pi i}{N}\right)\right]/4\right\} \cdot B_M \exp\left(\frac{2\pi i y}{MD}\right) \qquad (9)$$

Thus Eq. 9 indicates that the amplitude of the once-smoothed $\underline{x}$-feature has been replaced by $^1\bar{A}^x_N$

$$ {}^{1}\overline{A}_N^{x} = A_N \left(1 + \cos \frac{2\pi}{N}\right)\Big/2 \qquad (10) $$

that is, by $A_N R_N^{(1)}$ where $R_N^{(1)}$ is the response function for a single $\underline{x}$-smoothing on a wave feature of N grid lengths. The case $N = 1$ is excluded since at least three grid points are required to define the shortest wave-feature. For $N = 2$, $R_N^{(2)} = 0$, so that a single smoothing with 1 - 2 - 1 weighting factors completely filters waves of wavelength 2D, in both $\underline{x}$ and $\underline{y}$ directions. The results of eight consecutive such one-dimensional smoothings (where possible) results in a response factor $R_N^{(8)}$

$$ R_N^8 = \left[\left(\frac{1 + \cos \frac{2\pi}{N}}{2}\right)\right]^8 \qquad (11) $$

so that for $N = 4$, $R_4^{(8)} = 1/256$, while for $N = 8$, $R_8^{(8)} = .06254$. On the other hand, for $N = 24$, the amplitude is reduced only by the ratio indicated by the response factor $R_{24}^{(8)} = 0.7635$. Thus medium and long waves are altered only slightly, while mesoscale and very short synoptic waves are essentially filtered out. It was found by testing that radiometric channel which had the greater mesoscale detail (channel 5), that the criterion of eight smoothing passes in both the $\underline{x}$-and $\underline{y}$-directions produced the result

$$ \frac{\sigma^{(8)} - \sigma^{(7)}}{\sigma^{(0)}} \leq .001 $$

where $\sigma^{(0)}$ is the standard deviation of the original field $\tilde{T}_5$, and $\sigma^{(8)}$ is that of the eightfold smoothed $\tilde{T}_5$ field. Thus the use of 8-by-8 smoothing passes

in both the x - and y-directions was taken to be the desired degree of filtering for the "smoothed" variables of this study, both with respect to the contour-height and the MRIR variables for each day.

In the case of the contour fields, all of which have well-defined rectangular boundaries, the maximum number of 8 x-smoothing passes could only be performed where either I and/or 77-I exceeded 8. Where either I and/or 77-I were less than 8, only (I-1) or (77-I-1) x-smoothings were possible. A similar remark applies to the number of y-smoothings relative to the triad of test counters (J-1, 65-J-1, 8), the minimum value of the triad being chosen. Thus for example, on the boundaries I = 01 and I = 77 only y-smoothings up to the maximum of eight were allowed, whereas on J = 01 and J = 65, only x-smoothings up to the maximum were permitted. In order to permit the maximum of both 8 x- and y-smoothings a point in the contour field had to lie within an inner core displaced 8 grid spaces from the rectangular outer boundary of Fig. 1.

If the resultant smoothed field is denoted simply $Z(x, y)$ the original field $\tilde{Z}$ is simply expressible as

$$\tilde{Z}(x, y) = Z(x, y) + Z'(x, y) \tag{12}$$

where Z' is the residual height value at point (I, J). An IBM-360 computer program was designed to perform the smoothing operation from the original 65-by-77 card-punched contour fields at both 500-and 1000-mb, and then to store both the smoothed fields and residual fields in convenient locations on magnetic tape.

A mapping program using the same scale as Fig. 1 was developed to print both the smoothed fields and the residual fields. These fields were then carefully copied onto the matching geographic acetates of Fig. 1. The following examples for days 3 and 4 (Table 1) are shown in order as Figs. 2a and 2b, Fig. 3, Fig. 4, Fig. 5:

Fig. 2a, the smoothed 500-mb contour field for day 3

Fig. 2b, the residual 500-mb field for day 3

Fig. 3, the smoothed 1000-mb field for day 3

Fig. 4, the smoothed 500-mb field for day 4

Fig. 5, the smoothed 1000-mb field for day 4

Only one example of a residual contour field is shown since only statistical use is made of the smoothed variables in this study. All of the residual fields for all days of Table 1 have also been stored on magnetic tape so that later investigations involving their use may be carried out. In this study however, the extreme value of a 500-mb residual was 20 gpm, as compared with a typical 500-mb smoothed value of 5600 gpm, so that the decomposition process into a smoothed analysis reduces the detail of $\tilde{Z}_5$ only slightly. While the relative magnitudes of residual to smoothed variables for the other fields is more appreciable, inclusion of the residuals in the stepwise regression equations for Z_5 led, upon experimentation, to greater values of unexplained variance.

Smoothing of the MRIR fields. As noted before, the MRIR composite temperature fields were card-coded exactly in the same format used for the $\tilde{Z}_5$ and $\tilde{Z}_{10}$ fields. However, for each of the 4 MRIR days, there was a fairly extensive area

in the northwest section of Fig. 1, where zeros appeared on the NASA composite grid-print charts. A schematic version of the data-void section, and the adjacent body of data for all five MRIR channels on day one is shown in Fig. 6. The reason for the existence of the data-void is that the satellite data tapes for the final two orbits fall within command range of the Gilmore Creek data acquisition center, and pass over into readout mode. As long as the data-boundary progresses from lower left to upper right, with no kinks such as is depicted by the segment A E C D B of Fig. 6, only a slight modification of the procedure applied to the previously described rectangular arrays permits the smoothing to be applied to the MRIR fields. In order to apply the preceding smoothing procedure when a boundary kink such as A C B exists, the kink must be rectified, so as to substitute the revised boundary A B for A E C D B. This was programmed as follows: in the scanning of the card-deck by J-values when a segment of the boundary, such as C B, was found to retreat toward lower I-values, (so that zeros were actually coded at the grid positions within B C A), a subroutine "FILL" was employed. This program had the effect of inserting the value at D into the void between D and E, and the value at B into the 3 data-void positions between D and A. There were no kinks on days 2 and 3, but such a kink involving a total of 10 "fills" was encountered also on day 4. After rectifying all such kinks in the boundary to be parallel to the J-axis, the smoothing procedure outlined for rectangular arrays may be applied. The revised boundary, having no decreasing I-abscissa in proceeding from SW toward NE, is now treated as an unsmoothed (original) set of field values along its entire length. For an x-point lying one grid interval within the revised boundary, a single x-smoothing is permitted; and the smoothing count increased to a maximum of 8 x-smoothings at 8 or more grid lengths inward parallel to the I-axis.

For a y-point lying one grid interval below the revised boundary, a single y-smoothing is permitted, and in a similar way one proceeds to 8 y-smoothings at eight or more grid intervals below the revised data boundary shown in Figs. 6 and 9. The remaining MRIR boundaries, which are boundary-segments of the original 65-by-77 rectangle, are handled precisely as was done before for the purely rectangular case, discussed just above Eq. 12.

The error introduced in the "original" MRIR fields in kink-filled segments, such as shown in Fig. 6, tends to contribute some minor inconsistencies to the regressions near the boundary, but will have little influence on the 8-by-8 smoothed MRIR values within the main body of the data fields since the number of point values inserted for boundary-rectification were few in number. A somewhat more serious error lies in the use of the rectangular boundary for initiation of the contour-field smoothing, while using the rectified boundary for starting the MRIR-variable smoothing process.

Again, the decomposition suggested by

$$\tilde{T}_i (I, J, k) = T_i (I, J, k) + T'_i (I, J, k) \qquad (13)$$

$i = 1,2,3,4,5$ indicates channel number

$k = 1,2,3,4$ indicates day number (from Table 1)

was performed by the smoothing program applied with the rectified "northwest" boundary condition. The smoothed fields T_i were stored on the same magnetic tape as the Z_i-fields, ordered similarly by coordinate location (I , J) and by common day number. Residual fields were also stored in a similar manner, but the latter fields were not dealt with statistically in this paper.

In order to map the T_i fields, a specific multiple of ten degrees Kelvin, just lower than the lowest temperature in the body of the original $\tilde{T}_i$ field, was introduced into the area northwest of the revised boundary (see Fig. 6). This device permitted map-contouring with a minimal range of band-indexing digits for the original, smoothed and residual fields. Examples of the smoothed-field mappings of MRIR day 3 for channels 2 and 3 are shown in Figs. 7a and 8. In Fig. 7b, the channel 2 residual field for the same day is shown for illustrative purposes only. These analyses have not carried out in the data-void area to the left of the rectified boundary.

The channel 3 smoothed-temperature analysis exhibits an easterly "radiative thermal wind" field, with superimposed troughs and ridge features.

A preliminary comparison of the $T_2\,(I, J)$ field for MRIR day 3, (Fig. 7a) with that of the two corresponding contour fields for day 3, seems to indicate a greater similarity of feature with Z_{10} rather than with Z_5. However, from this point on we will rely on the results of statistical field-regression analysis for the specification of Z_5.

5. GEOGRAPHIC AREAS FOR APPLICATION OF THE REGRESSION ANALYSES

At all (I, J) points of the grid, there exist smoothed-field values of the variables.

$$(Z_5,\ Z_{10},\ T_1,\ T_2,\ T_3,\ T_4,\ T_5)_{I,J}$$

for each of the four days identified in Table 1. As indicated in Section 4, some of the (I, J) points in the northwest corner of the radiometric fields are devoid

of data, and here all the T_i (I, J) are now reset to zero, and such (I, J) points are eliminated from the regression analysis. Recall that in rectification of the MRIR boundaries, some possible discrepancies in a few selected MRIR grid-values have been introduced.

Because of the weak gradients of Z_5(I, J) in the range J = 01, ... , 07, (Fig. 9), no specification was attempted in this zonal band. However, the step-wise regression analysis was applied separately in the remaining areas identified by geographical code letters A, B, C, D, shown in Fig. 9. The areas are bounded as follows:

(1) areas A and D selected within J = 08 to J = 23 inclusive, and the A/D boundary at I = 31, is included in A;

(2) areas B and C selected within J = 24 to J = 65, inclusive, and the B/C boundary at I = 31, in included in B.

The significance of the intermediate zone boundary J = 23, corresponding to 31N latitude, is that it separates extratropical areas from subtropical latitudes. The west-east divider at I = 31 has been selected in order to separate geographical regions more likely to be affected by Pacific influences from those originating in the Gulf of Mexico and/or the Atlantic Ocean.

Both areas B and C, while contained within the geographical limits schematically indicated in Fig. 9, are subject to some point-deletions made when the rectified MRIR boundary is introduced. This latter "data" boundary slants across the northwest corners of areas B and C in a manner dependent upon the particular subtracks of the last two orbits falling within the MRIR day considered (Table 1).

If $Z_5\ (I\ ,\ J)$ is to be specified in terms of area-sample sets $(Z_{10},\ T_1\ ,\ \ldots\ ,\ T_5)_{IJ}$, using a multiple regression analysis, the total number N of the sample sets within each area of Fig. 9 must be known. The number N of (I, J) grid points within areas A, B, C, D, of Fig. 9 is listed by individual day, and also for pooled three-day periods (Table 2). In addition, the corresponding numbers n of NMC grid points contained in the four areas, reconverted to the original polar stereographic contour maps (with a spacing of 381 km at 60N latitude) have also been listed in Table 2. For purposes of computing regression equations based upon pooled three-day samples, the population counts have been added together, although this is not to imply that the $(Z_5,\ Z_{10},\ T_1\ ,\ \ldots\ \ldots\ ,\ T_5\ )_{IJ}$ sample-sets are not spatially correlated, nor serially correlated at 24-hour intervals.

The Miller stepwise, screening regression (1962) is to be employed to derive $Z_5\ (I\ ,\ J)$ in a best-fit form

$$Z_5\ (I,\ J) = \overline{A} + (A_{10}\ Z_{10} + A_1\ T_1 + A_2\ T_2 + A_3\ T_3 + A_4\ T_4 + A_5\ T_5)_{I,\ J} \qquad (14)$$

The essence of the Miller regression technique has been converted to an equivalent form by the University of California at Los Angeles Health Sciences Computing Facility (Dixon, 1966), and designated by the program name BMD 02R. By use of this program, equations of form (14) are to be derived for each area and day, and by time-pooled areas. Such results are to be shown in later tables of statistics, such as Tables 3, 4, and 5. In the derivation of these statistics, the sample sizes were taken identical to the numbers N of (I , J) sample sets for each stratification. In order to assess the significance of the sth variable added in the stepwise regression, one should know, however, the number d of such (I , J) sample sets which are independent of one another. While there is no

precise way of making an a priori statement of the number of degrees of freedom d, the assumption has been made that d is proportional to the number n of NMC grid points contained within the data sample. Hence in Table 2, there is listed by area and by day, or pooled-set of days, both the number N of (I, J) samples, and n of (i, j) samples, $(Z_5, Z_{10}, T_1, T_2, \ldots\ldots, T_5)_{ij}$. The use of n "relative" degrees of freedom is based upon the RMSE verification practice commonly used in NMC numerical analysis procedures, the results of which are given in Section 7.

6. SPECIFICATION RESULTS OF THE STEPWISE, SCREENING REGRESSION ANALYSIS BY AREAS

Table 3 is presented next to be illustrative of the ordering and screening results of the stepwise regression program BMD 02R applied to the data-sets by area and by day. The multiple correlation coefficient, R_s, upon entry of the sth predictor, has been computed using all N of the data samples $(Z_{10}, T_1, \ldots\ldots, T_5)_{IJ}$ in each area, where N is the (I, J) grid-point sample size of the areas listed in Table 2. If the notation of the regression equation is temporarily altered to the following form

$$Y_1 = A_1 + A_2 X_2 + A_3 X_3 + A_4 X_4 + \cdots\cdots \qquad (15)$$

the third-step multiple regression coefficient, is usually denoted in standards texts [Kenney and Keeping, 1951, p.346] as $r_{1.234}$, but is henceforth denoted here by the more compact notation R_3, etc. The magnitude of this statistic is given by

$$R_3 = \sqrt{1 - (1 - r_{12}^2)(1 - r_{13.2}^2)(1 - r_{14.23}^2)} \qquad (16)$$

Eq. 16 may be generalized to s predictors.

TABLE 2. Total number N of non-void data cases at (I, J) grid points within the areas depicted in Fig. 9, by day and pooled days, and the corresponding number n of NMC grid points for the same area-days.

Area	Day 1		Day 2		Day 3		Pooled days 1, 2, 3		Composite[5] B-C pooled days 1, 2, 3		Day 4		Composite B-C	
	N	n	N	n	N	n	N	n	N	n	N	n	N	n
A	496	93	496	93	496	93	1488	279	——		496	93		
B	948	83	856	77	880	80	2684	240	8395	640	998	87	2930	223
C	1932	136	1911	134	1868	130	5711	400			1932	136		
D	736	143	736	143	736	143	2208	429	——		736	143		

[5]Day 4 data has been reserved for test-verification only in the composite zones B, C.

In (16), the lower case r's are either the simple correlation coefficient, or the partial correlation coefficients of first order ($r_{13.2}$), or of second order ($r_{14.23}$), depending upon the subscripted notation. Table 3 shows how an ordering of the predictors can be made for each day, but there still remains the question of statistical significance of the sth predictor added. Miller (1962) has described a rigorous criterion making use of the F test, together with the exact knowledge of the number of degrees of freedom at the sth step. With regard to Table 3, the question of assessment of significance of the predictors added to the stepwise regression equations is postponed until the pooled three-day regressions of Table 4 have been examined.

In applying the Miller screening technique to the three-day pooled samples of Table 4, the total number N of sample-sets of column 5, Table 2, have been employed. Using such time-pooled sample data, the resulting regression should prove to be stable, if reliable regressions exist at all. For each area A, B, C, D the stepwise ordering of the variables within the array (Z_{10}, T_1, ... , T_5) has been selected upon the basis of the highest multiple correlation coefficient R_s relating Z_5 (I , J) to the parameters already introduced at the s th step. In Table 4, the cumulative percentages of explained variances, R_s^2, at the sth step have been included.

The specification of Z_5 (I , J) in areas A and D. In both areas A and D, comparison of the values of R_s in Table 4 with those at the sth step for the several days in Table 3, indicates a well-defined reduction of the former in comparison with the corresponding R_s of Table 3 (s = 1, . . . , 6). The shrinkage in R_s is due partly to the lack of homogeneity of the order of the predictors added

at successive steps, and partly to the marginal stability of the individual regression equations through days 1, 2, 3. In order to avoid specific reference to the number of degrees of freedom in these areas, the minimum requirement for any predictor accepted as significant in Tables 4A,D is that it contribute at least two percent to the cumulative explained variance upon entry. For areas A and D, the resulting specifications reduce to the two-predictor system of screened regression equations:

$$\text{(A)}\quad Z_5 = 5352.781 + 2.06007\, T_4 + .16784\, Z_{10}$$

$$\text{(D)}\quad Z_5 = 5718.875 + .37917\, Z_{10} + .45440\, T_2 \qquad (17)$$

Eqs. 17A, 17D account only for .4135, .6030, respectively, of the cumulative percentage explained variances of the 500-mb height fields. The percentages of specification just cited still leave standard errors of estimate $0.7558\,\sigma_5$ (A) and $0.6301\,\sigma_5$ (D), where σ_5 is the standard deviation of Z_5 (I, J). While these standard errors amount only to 21.282 and 12.153 gpm, respectively, in areas A and D, the fractional degree of specification of Z_5(I, J) appears too small to be usefully related to the MRIR variables T_4 (I, J) or T_2(I, J). In this connection, note for example, the lack of spatial variation in the Z_5 (I, J) field on day 3, Fig. 2a, across the zone A-D of Fig. 9, and then observe the contrast in this respect with channel 2 (T_2-field) in the same zone of Fig. 7a. A similar contrast in Z_5 (I, J) relative to T_4 (I, J) applies in this zone in view of the high simple correlation coefficient r (T_2, T_4) = .965 found between the pooled T_2-and T_4-fields in the composite zone A-D. Thus either of the two fields, T_2(I, J) or T_4 (I, J), seems to depict the broad-scale cloud features and their spatial variations in and near

TABLE 3. Regression statistics in the specification of $Z_5(I, J)$, by individual zone and day. N indicates the grid-point sample size, for the days listed under each area.

Day No.	Area A (N = 496,496,496)		Area D (N = 736,736,756)		Area B (N = 948,856,880)		Area C (N = 1932,1911,1880)	
	Vrbl entered upon step s	Multiple correl. coeff., step s	Vrbl entered upon step s	Multiple correl. coeff., step s	Vrbl entered upon step s	Multiple correl. coeff., step s	Vrbl entered upon step s	Multiple correl. coeff., step s
1	T_4	.7167	Z_{10}	.7675	T_3	.8204	T_3	.8860
	T_1	.7404	T_1	.8049	T_2	.9474	Z_{10}	.9460
	T_2	.7621	T_3	.8347	T_4	.9519	T_2	.9528
	Z_{10}	.7691	T_5	.8467	T_5	.9542	T_4	.9547
	T_3	.7928	T_4	.8481	Z_{10}	.9546	T_5	.9561
	T_5	.7937	T_2	.8547	T_1	.9547	T_1	.9569
	Std. dev. of Z_5, σ_5 = 22.675 gpm		σ_5 = 18.239 gpm		σ_5 = 94.588 gpm		σ_5 = 167.805 gpm	
	Std. error of est., S_5 = 13.877 gpm		S_5 = 9.507 gpm		S_5 = 28.232 gpm		S_5 = 48.807 gpm	
2	T_4	.7029	Z_{10}	.8945	T_3	.9036	T_3	.9302
	T_1	.7203	T_1	.9001	T_2	.9728	T_2	.9595
	T_3	.7235	T_2	.9090	T_1	.9741	Z_{10}	.9664
	Z_{10}	.7356	T_3	.9100	T_4	.9758	T_5	.9677
	T_5	.7421	T_5	.9112	T_5	.9760	T_1	.9685
	T_2	.7421	T_4	.9113	T_{10}	.9760	T_4	.9707
	Std. dev. of Z_5, σ_5 = 27.329 gpm		σ_5 = 19.354 gpm		σ_5 = 112.607 gpm		σ_5 = 190.367 gpm	
	Std. error of est., S_5 = 18.431 gpm		S_5 = 8.004 gpm		S_5 = 24.489 gpm		S_5 = 45.846 gpm	
3	T_1	.7532	T_3	.8025	T_3	.9487	Z_{10}	.8616
	T_5	.7739	T_1	.8111	T_2	.9662	T_3	.8899
	Z_{10}	.7878	T_4	.8447	T_5	.9688	T_2	.9109
	T_2	.7884	T_5	.8699	T_1	.9699	T_5	.9312
	T_3	.7921	Z_{10}	.8741	Z_{10}	.9705	T_4	.9489
	T_4	.7945	T_2	.8766	T_4	.9706	T_1	.9489
	Std. dev. σ_5 = 29.361 gpm		σ_5 = 17.779 gpm		σ_5 = 106.447 gpm		σ_5 = 195.956 gpm	
	Std. error of est., S_5 = 17.939 gpm		S_5 = 8.592 gpm		S_5 = 25.717 gpm		S_5 = 61.925 gpm	

TABLE 4: Summary of the stepwise regression analysis for Z_5 (I, J) for the three-day pooled samples corresponding to days 1, 2, 3 and the geographic areas A, B, C, D of Fig. 9. All statistics listed, except F-upon-entry, have been computed based upon N-sample-sets, whereas F-values are computed in B and C using only n as the number of degrees of freedom (see Table 2 for N, n).

Area and sample-size N	Predictor[6] selected at step s	sth predictor		Mult. correl. coeff., sth step	Cum. percent expl. variance, sth step	F-upon-entry, sth predictor	Coeff. of predictor in Eq. 14
		Mean	Std. dev.				
A (N = 1488)	T_4	263.808	7.596	.6071	.3684	Not computed in area A	2.06007
	Z_{10}	103.065	36.066	.6431	.4135		0.16784
	T_5	268.539	16.893	.6534	.4269		–
	T_3	224.184	1.307	.6594	.4347		–
	T_2	281.478	9.948	.6633	.4399		–
	T_1	233.343	7.931	.6738	.4540		–
predictand constant term, step 2	Z_5	5913.547	27.871				– 5352.781
D (N = 2208)	Z_{10}	153.747	35.945	.7628	.5819	Not computed in area D	0.37917
	T_2	282.458	6.405	.7765	.6030		0.45440
	T_1	231.406	5.056	.7838	.6144		–
	T_5	239.358	18.511	.7873	.6198		–
	T_4	263.608	5.485	.7877	.6205		–
	T_3	224.053	1.013	.7877	.6205		–
predictand constant term, step 2	Z_5	5905.523	19.291				– 5718.875
B (N = 2864)	T_3	228.878	1.995	.8746	.7649	774.335	-30.69087
	T_2	283.978	13.088	.9499	.9023	333.304	8.41798
	T_5	274.389	19.002	.9516	.9056	8.250	0.48098
	T_4	265.537	8.321	.9525	.9073	4.314	-6.40842
	Z_{10}	100.943	39.566	.9548	.9117	11.660	0.25845
	T_1	232.335	4.884	.9558	.9135	4.849	–
predictand constant term, step 5	Z_5	5807.845	104.857				– 11,983.926
C (N = 5711)	Z_{10}	86.214	81.580	.8804	.7715	1343.794	1.19119
	T_3	230.222	3.433	.9231	.8521	216.350	-20.28604
	T_2	271.515	11.645	.9340	.8723	62.641	7.87111
	T_5	270.205	34.886	.9398	.8833	37.232	0.86207
	T_4	257.892	7.241	.9419	.8871	13.261	-6.08460
	T_1	229.831	4.655	.9419	.8871	0.000	–
predictand constant term, step 5	Z_5	5807.845	104.857				– 9,438.535

[6]All units of T_i are in °K, while those of Z_{10}, Z_5 are in gpm.

TABLE 5. Summary of the stepwise regression for Z_5 (I, J) for the three-day pooled sample and the composite areas B and C. All statistics listed, except F upon entry, have been computed using N = 8395 (I, J) sample sets, whereas F-values are computed using only n = 640 relative degrees of freedom (see Table 2).

Predictor selected, step s	Predictor property		Mult. correl. coeff., sth entry	Cum. percent expl. variance sth entry	F_s upon sth predictor entry, using d = 640 in (18).	Coeff. of predictor in regression Eq. 14		
	Mean	Std. dev.				five-predictor case	four-predictor case	three-predictor case
T_3	229.645	3.129	.8584	.7369	1786.933	-23.54013	-25.79663	-28.86453
T_2	275.342	13.390	.9119	.8316	358.218	10.36465	5.25123	4.14898
Z_{10}	90.923	71.231	.9347	.8437	212.000	0.93728	0.91026	0.71506
T_5	271.660	30.775	.9410	.8855	65.441	0.89908	0.85917	–
T_4	260.185	8.356	.9445	.8921	38.780	-0.04288	–	–
Z_5 (predict-and)	5713.148	174.301	–	–	–	–	–	–
const. term in Eq. 14, five-predictor case						10028.360	–	–
const. term in Eq. 14, four-predictor case						–	9875.152	–
const. term in Eq. 14, three-predictor case						–	–	11134.332
Std. error of estimate, S_5 (I, J) at step 5 (in gpm)						57.282	–	–
Std. error of estimate at step s = 4, (in gpm)						–	59.005	–
Std. error of estimate at s = 3, (in gpm)						–	–	61.961

the Intertropical Convergence Zone across areas A-D. On the other hand, conventional 500-mb analyses do not inherently have this capability in this zone. Consequently Eqs. 17A,D were not used in a specification-application to the test data of day 4.

The regression analysis for areas B and C. The regression analyses of Table 3 for areas B and C give multiple correlation coefficients at the sixth and final predictor-entry which range from 0.95 to 0.97 during the three dependent days 1, 2, 3. Moreover, there is some day-to-day homogeneity in the order of entry of the predictors. In area B, the first two predictors to enter by the Miller stepwise regression procedure are consistently T_3 and T_2. The same is generally true of area C, except that the entry of Z_{10} alternates with T_3 and T_2 in the selection of the first three predictors accepted by the stepwise regression procedure. Thus there appears to be reasonable expectation for a stable three-day pooled regression equation in areas B and C. This is borne out by the results of Table 4, where the pooled multiple regression coefficients have undergone shrinkage of only approximately 0.02 at the final step, $s = 6$. The multiple correlation coefficient R_5 remains as high as 0.9558 in the case of pooled area B, while it is 0.9445 in area C, confirming the estimate of the stability of linear regressions with respect to the dependent samples.

In the final column of Table 4B,C the coefficients of the pooled forms of the five-predictor equations appear. The sixth variable listed in each case was T_1, and gives insignificant added specification to $Z_5(I, J)$, based upon the discussion below relationships (19). For purposes of inferring statistical significance at the sth step, the F_s statistic defined after Miller (1962), at 1 and $d - s - 1$ degrees of freedom, was used:

$$F_{\underline{s}}(1, \underline{d}-\underline{s}-1) = \left[\frac{(\text{Cum. \% exp. var., step } \underline{s}) - (\text{Cum. \% exp. var., step } (\underline{s}-1))}{1 - (\text{Cum. \% exp. var., step } \underline{s})}\right] \quad (18)$$

Here $\underline{d}$ is the number of degrees of freedom, that is the number of randomly selected samples contained in the original N sample sets. Because of spatial dependency, an estimate for $\underline{d}$ is necessary. This estimate, in the case of areas B and C, was taken equal to the number n of NMC grid points already listed in Table 2. In the pooled samples used for the deduction of $F_{\underline{s}}$ in Table 4, these n -values were counted and found to be 240 and 400, respectively. The rationale of adding daily n -counts to arrive at three-day pooled sample counts has already been discussed near the end of Section 5, in connection with Table 2, since the independent test is to be based upon NMC grid-point verification, using 24-hour serially related dependent time samples. The details of the significance test is given below. For the present, it is to be emphasized that $\underline{d}$ is not assumed equal to n , but choice of the latter number seems to give the proper selection and ordering of the variables both in Tables 4 (dependent-data) and Table 5 (independent data).

If $\underline{d}$ is approximated by n in Table 4B, 4C only for computing tentative F-values, one may reject the null hypothesis of an insignificant contribution to explained variance provided

$$F_{\underline{s}}(1, n - \underline{s} - 1) > F^{*}_{\underline{\alpha}/\underline{s}}(1, n - \underline{s} - 1) \geq 7.08 \quad (19)$$

Here the F^*-value is Miller's critical value (1962) designed to insure that the probability of accepting the null hypothesis when it is false (a type II error) will be less than 0.05. Note however that each $F_s \geq 7.08$ will then correspond to a type I confidence requirement of (1-.05/6) = 0.9927, a rather stringent confidence limit for a type I error test.

Examination of the second last column of Table 4B, with the value of d assumed to be n leads to the conclusion that T_4 and T_1 are unacceptable as predictors in area B, at the required joint type I and II confidence levels just set forth. However, the same column of Table 4C indicates that all variables in the set (Z_{10}, T_3, T_2, T_5, T_4) are acceptable as predictors at the indicated confidence levels. In both areas B and C, the variable T_1 was the last selected and failed to satisfy the significance test of (19); hence T_1 was rejected from future considerations as a possible predictor. Thus the ordered sets of predictors for Z_5 accepted in Tables 4B, 4C, were

(B) T_3, T_2, T_5, (T_4 doubtful[7]), Z_{10}

(C) Z_{10}, T_3, T_2, T_5, T_4

Note the inversion of Z_{10} in the ordering for areas B and C. Statistically, this is an expression of opposing magnitudes of the linear correlations r (Z_5, Z_{10}) in the two areas, e.g.,

$$r_B (Z_5, Z_{10}) = 0.114 \quad \text{while} \quad r_C (Z_5, Z_{10}) = 0.8804$$

[7]The variable T_4 could have been deleted but is tentatively being retained for the stepwise screening regression to be applied to the composite area-sample B – C (see Table 5).

Physically these last results mean that in area C, the fields of Z_5 and Z_{10} were very nearly in phase, possibly due to the greater reliance placed on Z_{10} in arriving at Z_5-values in region C, an area of comparatively sparse data. On the other hand, in area B the extrapolative procedure used in reduction of terrain pressure to sea-level tends to weaken the resulting simple correlation coefficient between Z_5 and Z_{10}.

Because of the near equality of the R_5 values for areas B and C (0.9558 and 0.9419, respectively), it was decided to perform the stepwise regression analysis upon the B-C composite areas for the pooled three-day period. In order to arrive at a solution to the dubious roles played by Z_{10} and T_4, best fit five-predictor, four-predictor and three-predictor regression equations are to be sought using the Miller stepwise, screening procedure. For the composite B-C sample, the results obtained are listed in Table 5. As expected, the multiple correlation coefficient, R_5 (B, C) is 0.9445 at step 5, which is merely the composite result of the individual values for areas B and C, making due allowance for the standard errors of estimate at step 5 in the two areas. (These standard errors were S_5 (B) = 31.20 gpm and S_5 (C) = 62.00 gpm.)

For the composite B-C sample, the Miller stepwise regression procedure led to the revised order

$$T_3,\ T_2,\ Z_{10},\ T_5,\ T_4$$

with each predictor now significant according to the F-test of (18, 19), with $\underline{s}$ = 1, ... , 5, compared to Miller's critical $F^*_{\alpha/\underline{s}}$ (1,634) = 6.68 (using $\underline{d}$ = 640 and $\underline{s}_{max}$ = 5).

In Table 5, note that the variable T_4, which now seems to be significant, is the last predictor to enter. Since the exact number of degrees of freedom is still in question, the test of optimum specification in this experiment will be based upon the best specification afforded by application of the three prediction equations of Table 5 tested against the independent test-data of day 4. Here the possibility noted previously by Lorenz (1956), that the most detailed regression equation may contain a higher noise-level when applied to independent data, will be investigated. This is a distinct possibility in regard to the predictor T_4 in view of the simple correlation coefficient r $(T_2, T_4) = 0.972$ found in area B-C for the time-pooled dependent data test.

7. TEST FOR OPTIMUM SPECIFICATION IN THE COMPOSITE AREA B-C

The three dependent regression equations arrived at in Table 5 may be represented in the matrix form

$$\hat{Z}_5 (I, J) = (a_0, a_1, a_2, a_3, a_4, a_5) \begin{pmatrix} 1 \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{pmatrix} \qquad (20)$$

where $a_4 = a_5 = 0$ in the three-predictor case, etc. The coefficients for the three prediction equations are listed in the final column of Table 5, and the ordered predictors $X_1, \ldots, X_5$ are found in column 1 of Table 5. Our immediate objective is to test the three regression equations

(a) three-predictor, $\hat{Z}_5 = 11134.332 - 28.86453\ T_3 + 4.14898\ T_2 + .71506\ Z_{10}$

(b) four-predictor, $\hat{Z}_5 = 9877.152 - 25.79663\ T_3 + 5.25123\ T_2 + .91026\ Z_{10} + .85917\ T_5$ (21)

(c) five-predictor, $\hat{Z}_5 = 10{,}028.360 - 23.54013\ T_3 + 10.36465\ T_2 + .93728\ Z_{10} + .89908\ T_5 - 8.04288\ T_4$

for optimum specification of the field of Z_5 (I, J) of day 4. To do this the sample set matrix $(T_3, T_2, Z_{10}, T_5, T_4)_{IJ}$ is introduced at each of the N = 2930 grid points in the composite area B-C of test day 4, so that the resulting $\hat{Z}_5$ (I, J) may be computed at each grid point by each of (21a, b, c). The resulting standard errors of estimate are then determined from the equation

$$S_5(N) \equiv \sqrt{\frac{(\hat{Z}_5 - Z_5)^2_{IJ}}{N - 1}} \qquad (22)$$

according as s_{max} = 3, 4, or 5 in Table 5. The smallest standard error occurred for the four-predictor case (Table 6). The question of optimum results are discussed in more detail when a comparison test using the NMC grid-point values has been described.

Specification of $\hat{Z}_5$ (i, j) at NMC grid points for day 4. For the independent test (day 4) applied to the specification of $\hat{Z}_5$ (i, j) at the NMC grid points contained within area B-C, it was necessary first to perform the interpolation of each field T_3, T_2, T_5, T_4 (I, J) to interpolated values in terms of NMC (i, j),

coordinates by the inverse interpolation scheme described just below Eq. 3. This was simply done by expressing (I , J) in terms of (i , j), using the geometric relationships, Eqs. 1, 2, 3, followed by the double Bessel interpolation program previously employed in Section 2. The values of Z_5 and Z_{10} (i , j) were known at NMC grid points without the necessity of interpolation. Eqs. 21a, b, c were applied to the n = 223 field-values of T_3, T_2, Z_{10}, T_5, T_4 within area B-C, so that $\hat{Z}_5$ (i , j) was computed by each of the three prediction equations of (21). For each prediction case, the root mean square error was computed by the analog of Eq. 22, i.e., by

$$S_5(n) = \sqrt{\frac{(\hat{Z}_5 - Z_5)^2_{ij}}{n-1}}, \quad n = 223 \tag{23}$$

There is no a priori assumption here that the same multiple correlation coefficient R_s of Table 5 was applicable to the smaller set of NMC grid points lying within B-C on day 4.

The standard errors of estimate S_5 (N), and root mean square errors, S_5 (n), of the three prediction equations of (21), have been computed for the two populations defined at (I , J) and (i , j) on day 4, and are listed in Table 6. In addition, the percentage explained variances computed from $1 - S_5^2/\sigma_5^2$ are listed for both populations in Table 6. In this connection, a separate F test on the standard-deviations squared led to the result

$$F = \frac{\sigma_5^2\,(N)}{\sigma_5^2\,(n)} = \left(\frac{173.145}{172.986}\right)^2 = 1.00184$$

indicating no significant differences existed in the variances of the Z_5 (I , J) and Z_5 (i , j) fields for day 4.

TABLE 6. Comparison of the specifications by the three regression equations 21a, b, c applied to test sample (day 4), for the (I, J) and (i, j) sample sets, comprising N = 2390 and n = 223 grid points, respectively.

Grid Employed	Percent. expl. var., $1 - (S_5/\sigma_5)^2$			Coeff. of determination			Std. error of est. (gpm)		
	by (a)	by (b)	by (c)	by (a)	by (b)	by (c)	by (a)	by (b)	by (c)
(I, J)	.7863	.8915	.8179	.8868	.9442	.9044	80.05	57.04	73.09
(i, j)	.7789	.8867	.8172	.8826	.9416	.9040	81.34	58.23	73.96

Table 6 shows that the four-predictor equation leads to the largest percentage of cumulative explained variance, and accordingly to the smallest standard error of estimate for the independent test data. This is true for both the (I, J) grid-sample sets, and also for the NMC grid sample sets. The comparative results for the four-predictor versus the five-predictor equation tend to confirm the comment expressed in the last paragraph of Section 6, regarding the redundancy of T_4 applied to both the dependent and independent data samples. On the other hand, deletion of T_5 results in a sizable reduction in explained variance, indicating that T_5 is useful in specification of Z_5 (I, J) or Z_5 (i, j), even after the prior inclusion of T_3, T_2, Z_{10}.

Perhaps the most remarkable feature of the comparison of the results of Table 6 with those of Table 5 lies in the stability of the specification statistics. The four-predictor equation applied to the independent data actually leads to a somewhat smaller standard error of estimate for $\hat{Z}_5$ (I, J) than that based upon the three-day pooled dependent B-C sample. The corresponding results obtained by (21b) using the 223 NMC grid-point sample gives only a slightly smaller percentage explained variance than that from the (I, J) sample sets, upon which Eq. 21b was derived. Use of an F test upon the ratio of the percentage explained variances of Table 6 indicates that the null hypothesis $1 - S_5^2(N)/\sigma_5^2(N) \neq 1 - S_5^2(n)/\sigma_5^2(n)$ must be rejected at confidence level far in excess of 99%. This in turn indicates that in the regression analyses applied to the two different sets of grid-sample data of day 4, the number of sample-set replicates contained within the larger (I, J) population appears to be of the order of 2930/223. While no definitive statement may be made regarding the number of spatially independent

degrees of freedom on a given day, this question becomes an academic one since the four-predictor equation gives highly significant specification at NMC grid points. It should be recognized, however, that there may be a substantial serial correlation, since the independent test was made for 24-hours after day 3 in the series of 24-hourly dependent data. The results, however, indicate the feasibility of making 500-mb height analyses using MRIR data when sypplemented by an analysis of the Z_{10}-field as an "air-mass conditioner." The Z_{10}-field may well have been taken as that available 12-hours prior to verification of the Z_5-field, although experimental verification of this suggestion was not attempted in this study.

8. COMPARISON OF THE PREDICTED AND VERIFYING MAPS IN AREA B-C

The $\hat{Z}(I, J)$ analysis in area B-C for day 4 was computed using the four-predictor equation 21b and the corresponding day 4 predictor-variable matrix. The resulting map, $\hat{Z}_5(I, J)$ is shown in Fig. 10 for areas B-C. This map may be visually compared with the verifying FNWF map (see Fig. 4), poleward of J = 23 (latitude 32N), except in the MRIR data-void area. The derived pattern of Fig. 10 captures the longer wave features of Fig. 4, e.g., the long-wave ridge in western Canada, and the long-wave trough located along a line from north central Labrador through James Bay.

With regard to the short-wave features in Fig. 4, the derived $\hat{Z}_5(I, J)$ over-accentuates the short-wave trough near Newfoundland, by specifying a 5400 gpm closed low to the northeast of Newfoundland. A second short-wave trough in Fig. 4 seems to be passing through the major trough referred to in the preceding paragraph, and in Fig. 10, this is overaccentuated and also advanced too far to

the east. A third short-wave trough appears in Fig. 10 in a line from Minnesota to James Bay, and this feature has evidently been smoothed out in the official analysis in Fig. 4, although there is evidence of an east-west oriented 1000-mb trough somewhat north of James Bay. Each one of the predicted troughs in $\hat{Z}_5(I, J)$ referred to is strongly correlated with the coexistence of low values in the $T_2(I, J)$ field (Fig. 7A) for day 3, which by day 4 have moved out to precede the $Z_5(I, J)$ short-wave troughs.

Fig. 11 shows the error field $\epsilon(I, J) = \hat{Z}_5 - Z_5$, in the area B-C for day 4. Largest negative errors again turn out to be in near coincidence with low-values in $T_2(I, J)$, allowing for the advance of T_2-features between days 3 and 4. The root mean square error in Fig. 11 is 57.04 gpm, as indicated by Table 6. Part of this error is due to the boundary errors along the northwest MRIR data-boundary, where the smoothing procedures for the contour-height and for MRIR variables were of necessity somewhat different (Section 4).

9. CONCLUDING REMARKS

While the root mean square error remaining after the regression analysis appears to be large, a verification scoring system based upon the gradient of contour height would appear less severe. The regression-generated map (Fig. 10) had considerably more detail of short-wave feature than was indicated on the map used for verification, which was evidently more heavily smoothed in data-processing. The "predicted" $Z_5(I, J)$ short-wave features appear to be highly correlated with channel 2 wave-features, and secondarily with channel 5 features. One is led to raise the question of the possibility of oversmoothing in the conventional analysis procedures now in use, and whether a smaller grid-mesh and

a smaller limiting radial scan area per grid point might not lead to analyses more nearly compatible with satellite data.

The work of Smith (1968) using simulated (computed) values of the expected readouts from the Satellite Infrared Radiation Spectrometer (SIRS) may be regarded as a sophisticated analog of the work presented here. Smith selected seven narrow spectral intervals in the carbon dioxide band, and one in the window region near 11 microns, and ascribes "brightness" temperatures, and in turn, ambient temperatures, to thin layers scanned by his SIRS simulated instrument. It is noteworthy that the results of the present study using the NIMBUS II infrared and solar-reflectance radiances gave its maximum information in the same general spectral regions, channels 3 and 2 (followed by channel 5, which afforded some additional information regarding the cloud-cover aspect). Also independently, both this work and Smith's have made use of 1000-mb heights as a lower boundary in the regression analyses performed.

A comparison of the present method of height-specification with that of SIRS, based upon the NIMBUS 2B satellite data, when available in the mid-1969, is highly desirable, especially for a winter situation.

ACKNOWLEDGMENTS

I am deeply grateful to Mr. W.R. Bandeen, of the Planetary Radiations Branch, Goddard Space Flight Center, for stimulating discussions regarding numerous aspects of this work. Appreciation is expressed to Messrs. Ron Brunell and William Ehrman of the Navy Postgraduate School Computer Facility, both of whose ingenious programming contributions made the data reduction possible. Appreciation is also due to Mr. Ken Roberts for his careful drafting of the figures.

REFERENCES

Dixon, W. J., (Ed.), 1966: Biomedical Computer Programs, Health Sciences Computing Facility, University of California at Los Angeles, 585 pp.

Jensen, C. E., J. S. Winston, and V. R. Taylor, 1966: 500-Mb heights as a linear function of satellite infrared radiation. Mon. Wea. Rev., 94, 641-649.

Kenney, J. F., and E. S. Keeping, 1951; Mathematics of Statistics (part two), Princeton, N.J., D. Van Nostrand Company, 429 pp.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Scientific Report No. 1, Statistical Forecasting Project, Dept. of Meteorology, Massachusetts Institute of Technology, Cambridge, Mass., 1956, 49 pp.

Martin, F. L. and G. Warnecke, 1967: Some statistically-derived relationships involving medium-resolution radiation and conventionally-analyzed data fields. (Unpublished technical memo., copies available upon request from first-named author, 25 pp).

Miller, R. G., 1962: Statistical prediction by discriminant analysis. Meteor. Monogr., 4, No. 25, Boston, Mass., Amer. Meteor. Soc., 53 pp.

Ruff, I., R. Koffler, S. Fritz, J. S. Winston, and P. K. Rao, 1968: Angular distribution of solar radiation reflected from clouds as determined from TIROS IV radiometer measurements. J. Atmos. Sci., 25, 323-332.

Schuman, F. G., 1957: Numerical methods in weather prediction: II. Smoothing and filtering. Mon. Wea. Rev., 85, 357-361.

Smith, W . L., 1968: Statistical estimation of the atmosphere's pressure-height distribution from satellite radiation measurements. Paper presented at the IUGG/WMO Symposium on Radiation, Including Satellite Techniques, Bergen, Norway, Aug. 1968. (To be published as an ESSA Technical Report, NESC-48.)

Staff Members, 1963: TIROS IV Radiation Data Catalog and Users' Manual, National Space Science Data Center, Goddard Space Flight Center, Greenbelt, Md., 250 pp.

Staff Members, 1966a: NIMBUS II Data Catalog, Vol. 2, National Space Science Data Center, Goddard Space Flight Center, Greenbelt, Md., 297 pp.

Staff Members, 1966b: NIMBUS II Users' Guide, National Space Science Data Center, Goddard Space Flight Center, Greenbelt, Md., 229 pp.

APPENDIX

Conversion From (I, J) Map-Coordinates to Coordinates of Longitude and Latitude

In Fig. 1, the longitude λ at any grid-point abscissa I is given by

$$\lambda = 135 - (I - 1)(1.25) \qquad I = 1, \ldots\ldots, 77, \tag{A-1}$$

in degrees west longitude. The corresponding mapped distance ΔX at latitude φ for unit map increment $|\Delta\lambda| = 1.25$ degrees longitude is

$$\Delta X = a \left(\frac{1.25\pi}{180}\right) \cos\varphi_0. \tag{A-2}$$

where $\varphi_0 = 22.5N$ is the latitude at which the Mercator mapping is true, and a is the mean radius of earth (assumed spherical).

Since the northward grid interval ΔY is equal to ΔX, both represented by 1/2 inch, and also since ΔY must be true at $\varphi = \varphi_0$, one may write

$$\Delta Y = a \sec\varphi \cos\varphi_0 \Delta\varphi \tag{A-3}$$

and therefore (after replacement of finite increments by differentials)

$$\sec\varphi \, d\varphi = \left(\frac{1.25\pi}{180} \cos\varphi_0\right) dJ \tag{A-4}$$

$$J = 1, \ldots\ldots\ldots, 65$$

In (A-4), dJ represents the incremental counting number used to express northward map distances in terms of grid intervals (dJ = dI).

Integration of (A-4) from $J_0 = 15.4753$ corresponding to 22.5N latitude, to any allowable J within the range $\varphi_1 = 5N$ to $\varphi_2 = 65N$, leads to the result

$$\ln\left[\frac{(1+\sin\varphi)/\cos\varphi}{(1+\sin\varphi_0)/\cos\varphi_0}\right] = .021817(J - J_0) \qquad \text{(A-5)}$$

Inserting the functional values for $\sin\varphi_0$, $\cos\varphi_0$, followed by exponentiation, yields the expression for $\sin\varphi$

$$\sin\varphi = \left\{\frac{2.23985\left[\exp\left(.043634(J-J_0)\right)\right] - 1}{2.23985\left[\exp\left(.043634(J-J_0)\right)\right] + 1}\right\} \qquad \text{(A-6)}$$

for all values of $J = J(\varphi)$, starting $J = 1$ at $\varphi = 5N$.

The desired map-transformation from (I, J) to (λ, φ) coordinates is given by Eqs. A-1, A-6.

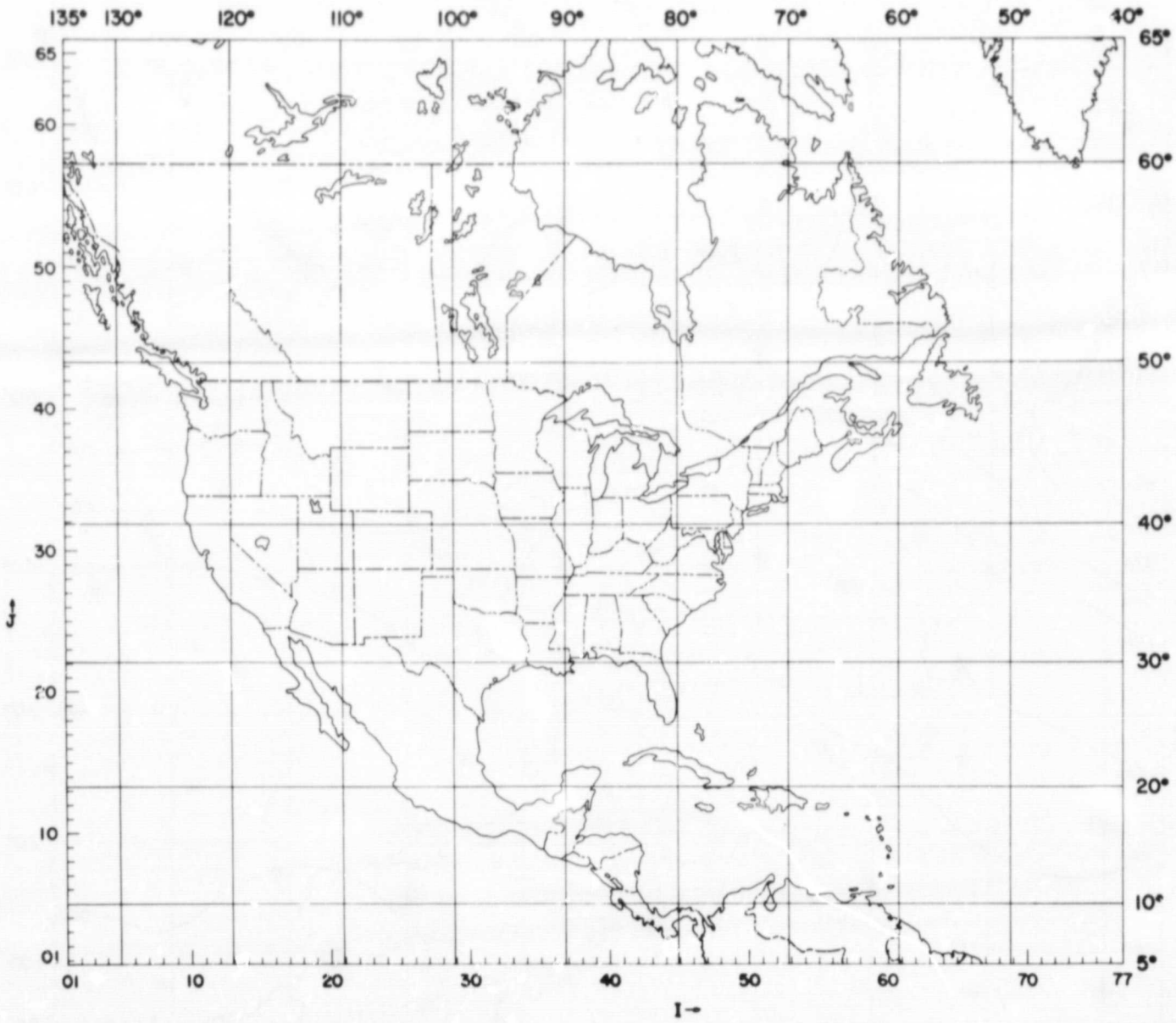

Figure 1—Mercator map-projection of data region used in analysis of radiometric and contour-height fields considered in this study. Equally spaced I- and J-intervals are indicated along the bottom and left boundaries of the rectangular area. Corresponding longitudes and latitudes are shown along the opposite boundaries.

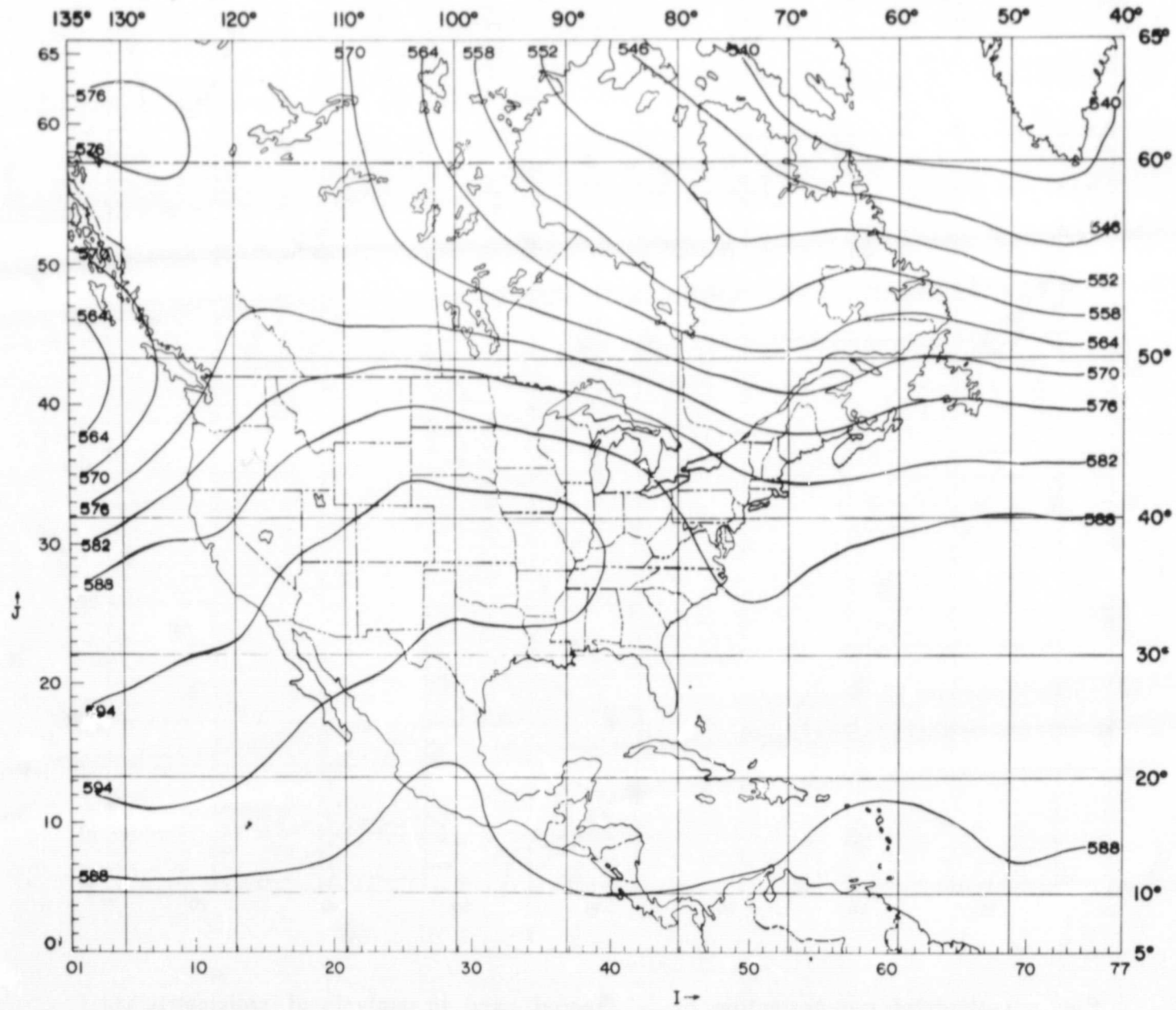

Figure 2a—The smoothed field of 500-mb contour heights $Z_5(I, J)$ for 00GMT, 18 July 1966, in units of geopotential decameters.

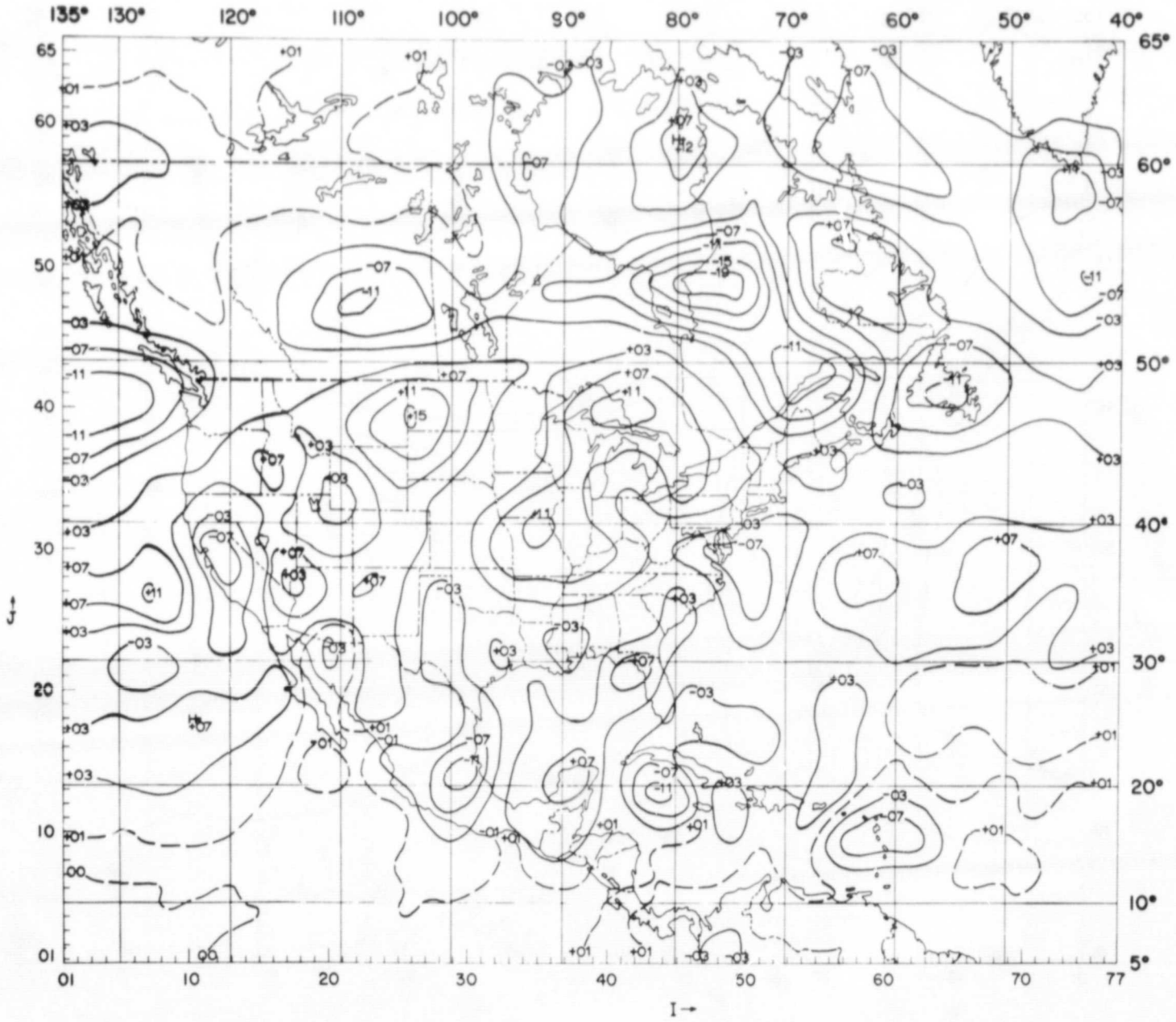

Figure 2b–The 500-mb residual field $Z_5'(I, J)$ in units of gpm for the same time and date corresponding to Figure 2a. The sum of Z_5' and Z_5 at each (I, J) gives the original 500-mb height field before smoothing was applied.

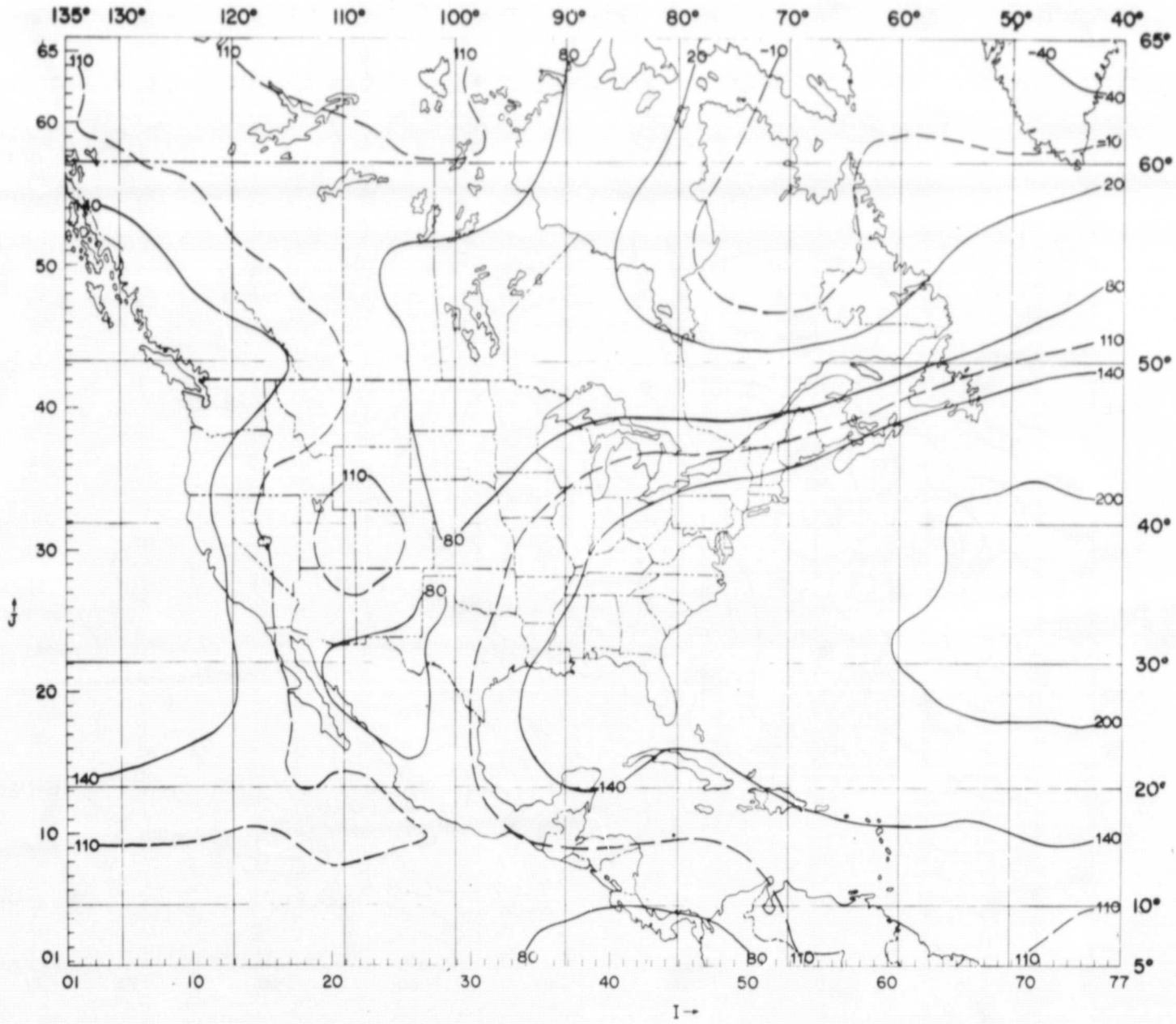

Figure 3–The smoothed field of 1000-mb contour heights $Z_{10}(I, J)$ of 00GMT, 18 July 1966, in units of gpm.

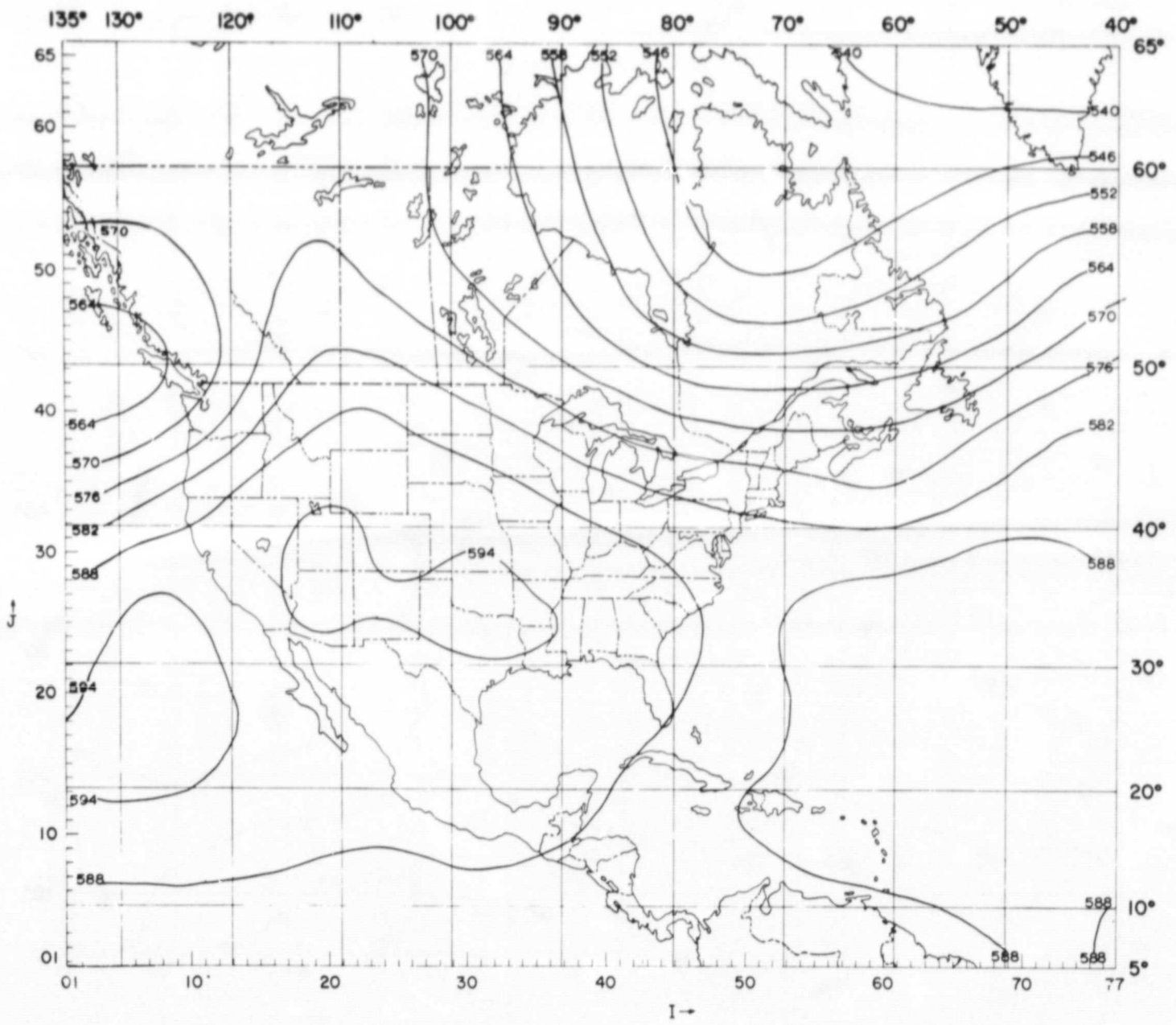

Figure 4—The smoothed field of 500-mb contour heights $Z_5(I, J)$ for 00GMT, 19 July 1966, in units of geopotential decameters.

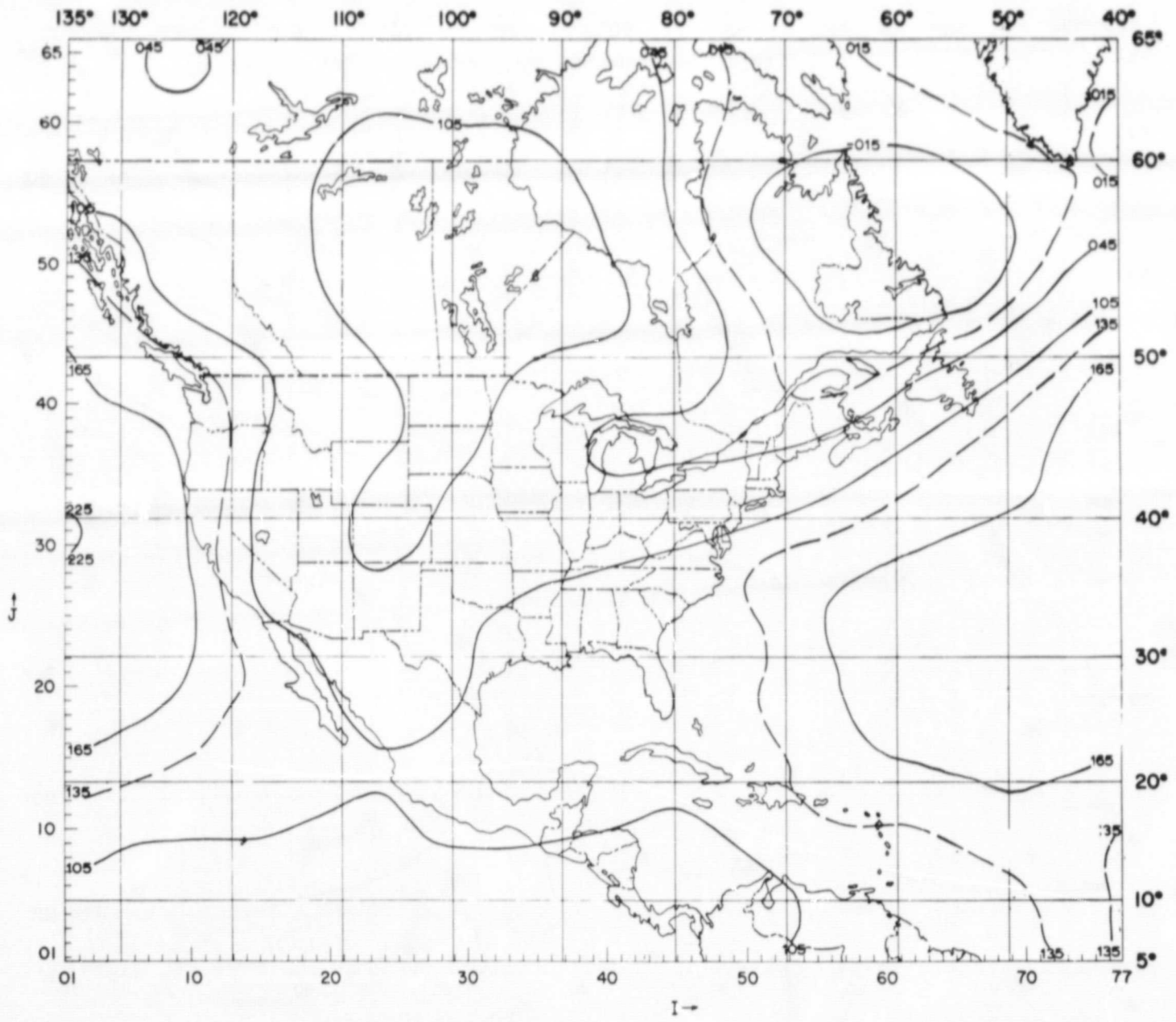

Figure 5–The smoothed field of 1000-mb contour heights $Z_{10}(I, J)$ for 00GMT, 19 July 1966, in units of gpm.

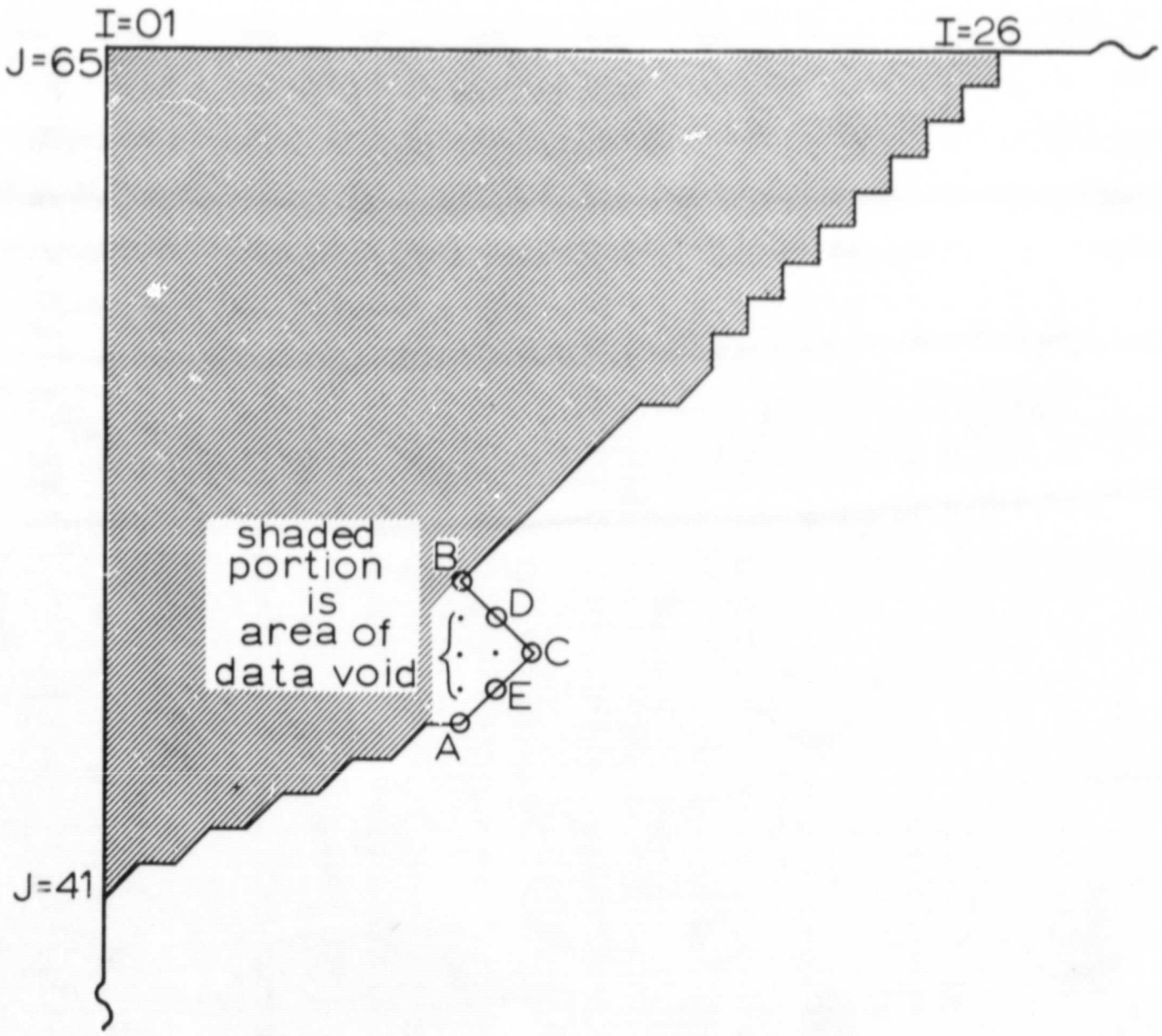

Figure 6–Schematic version of the northwest portion of a medium resolution radiometric field showing a typical data-void area (shaded portion), and a boundary kink indicated by the set of data-points AECDB. For application of the smoothing technique described in Section 4, the boundary kink is rectified by replacing the boundary AECDB by the line AB parallel to the J-axis.

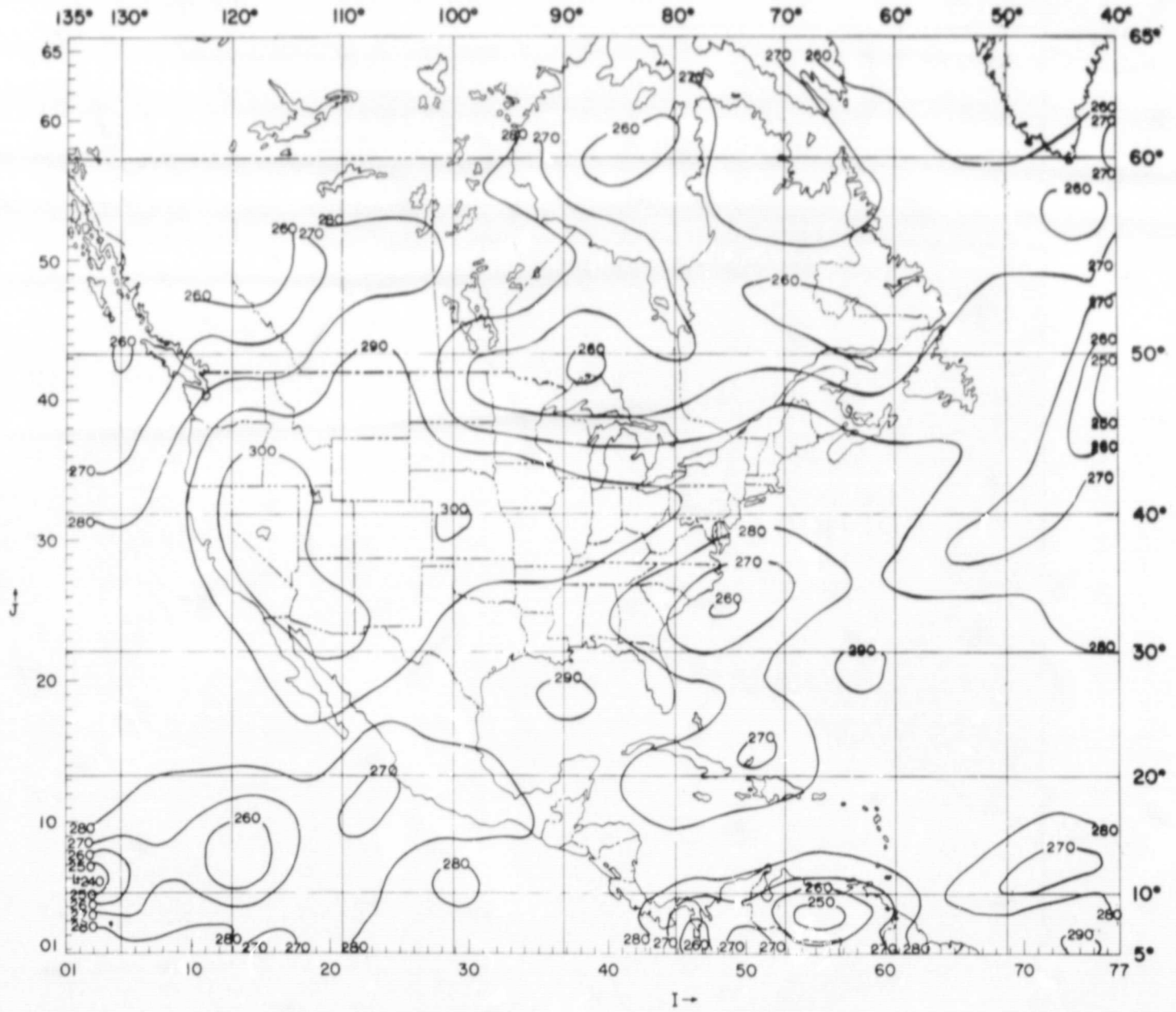

Figure 7a—The eightfold-smoothed map version of channel 2 equivalent blackbody temperatures in °K for MRIR day 17 July 1966 (day three), over the region of composited data-coverage. Isotherms are drawn at intervals of 10°K.

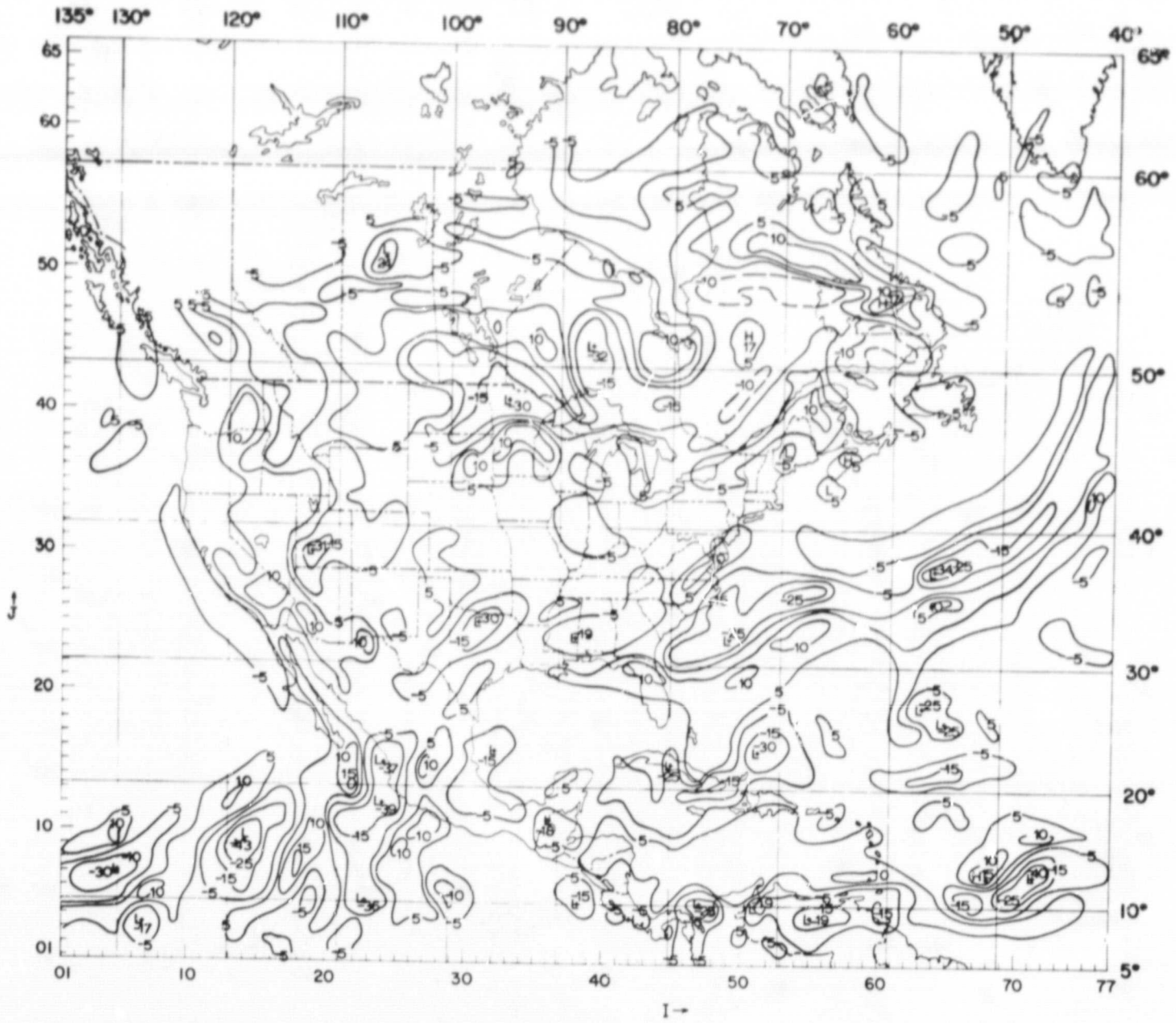

Figure 7b–The corresponding channel 2 residuals in °K for the same data period as in Figure 7a, with isotherms drawn at 5°K intervals, space permitting. The zero isotherm is omitted. Values of residuals at centers, with appropriate signs are indicated.

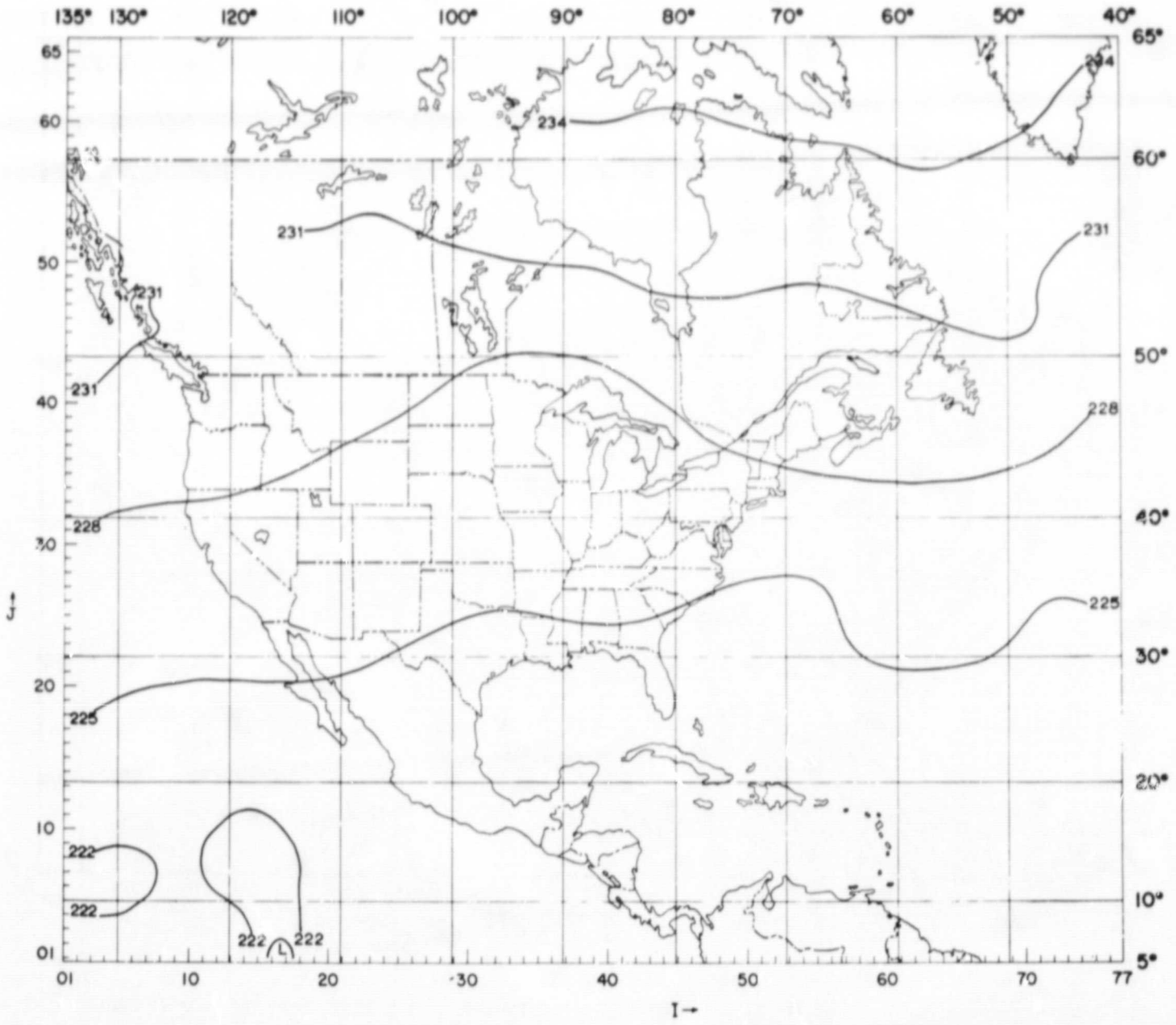

Figure 8–The eightfold-smoothed map version of T_{BB} for the composited field of channel 3 in °K for MRIR day 17 July 1966 (day three) over the region of MRIR data-coverage. Isotherms are drawn at intervals of 3°K.

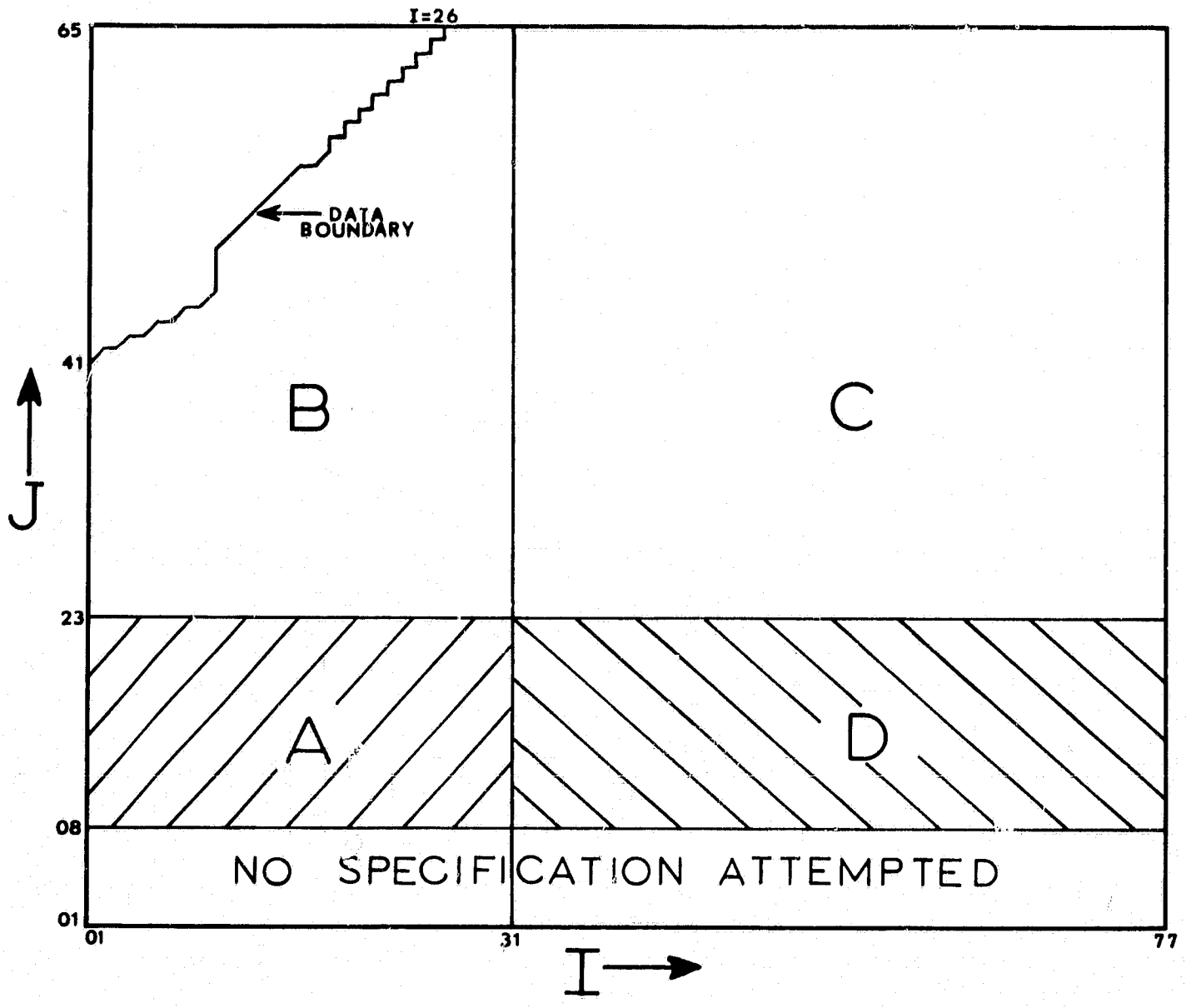

Figure 9–Schematic illustration of the stratified areas where the stepwise regression analysis was performed. The vertical line I = 31 indicates the eastern boundary of areas A, B. The horizontal lines J = 08 and J = 23 denote latitudinal boundaries for areas A and D. Note the data-gap in the northwest sector of area B which effectively delineates the northwest data-boundary of area B.

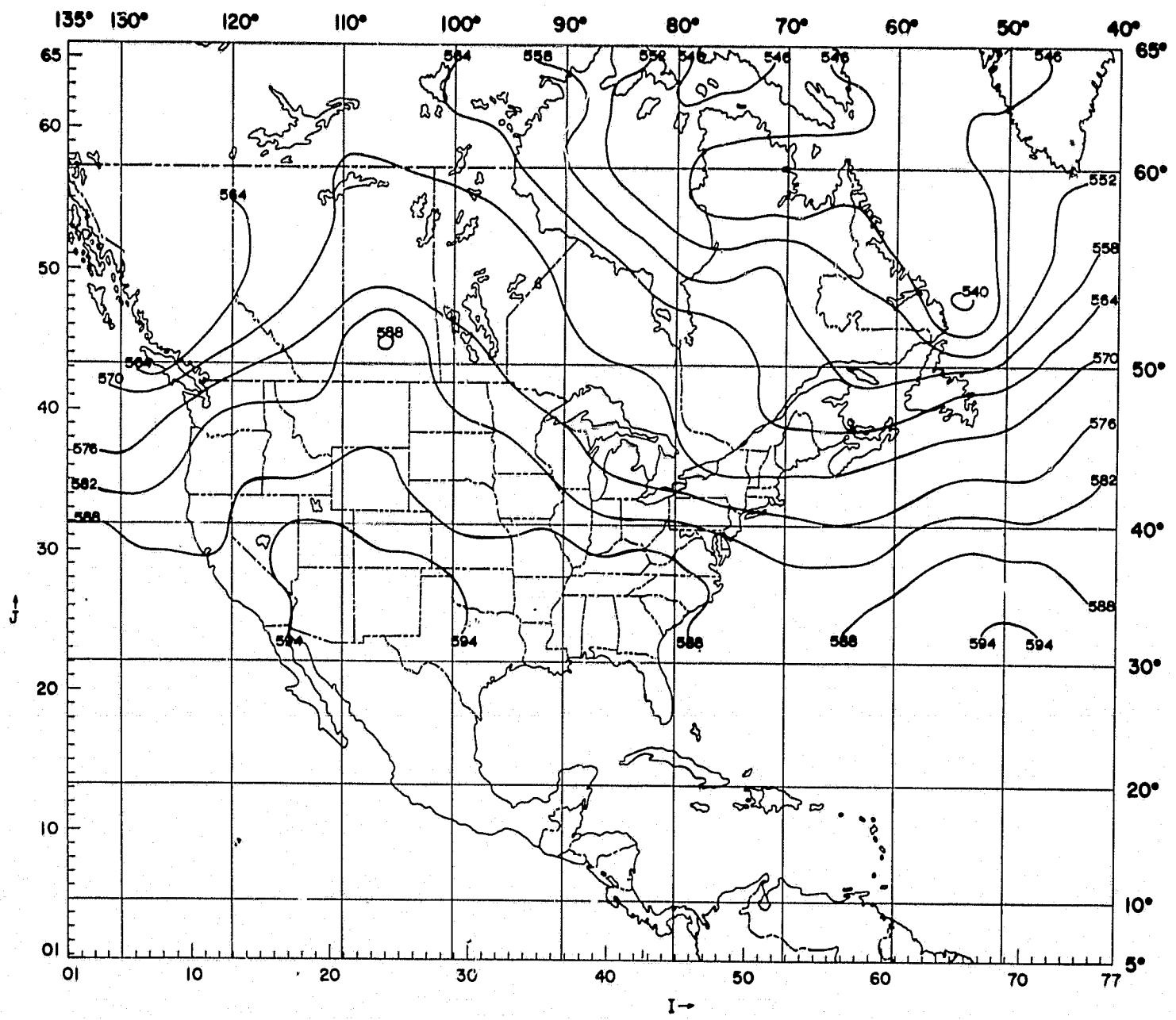

Figure 10–The $\hat{Z}_5(I, J)$ specification field for the composite area B–C on day 4 (00GMT, 19 July) for the independent data-case, using the four-predictor regression equation, which gave optimum verification results.

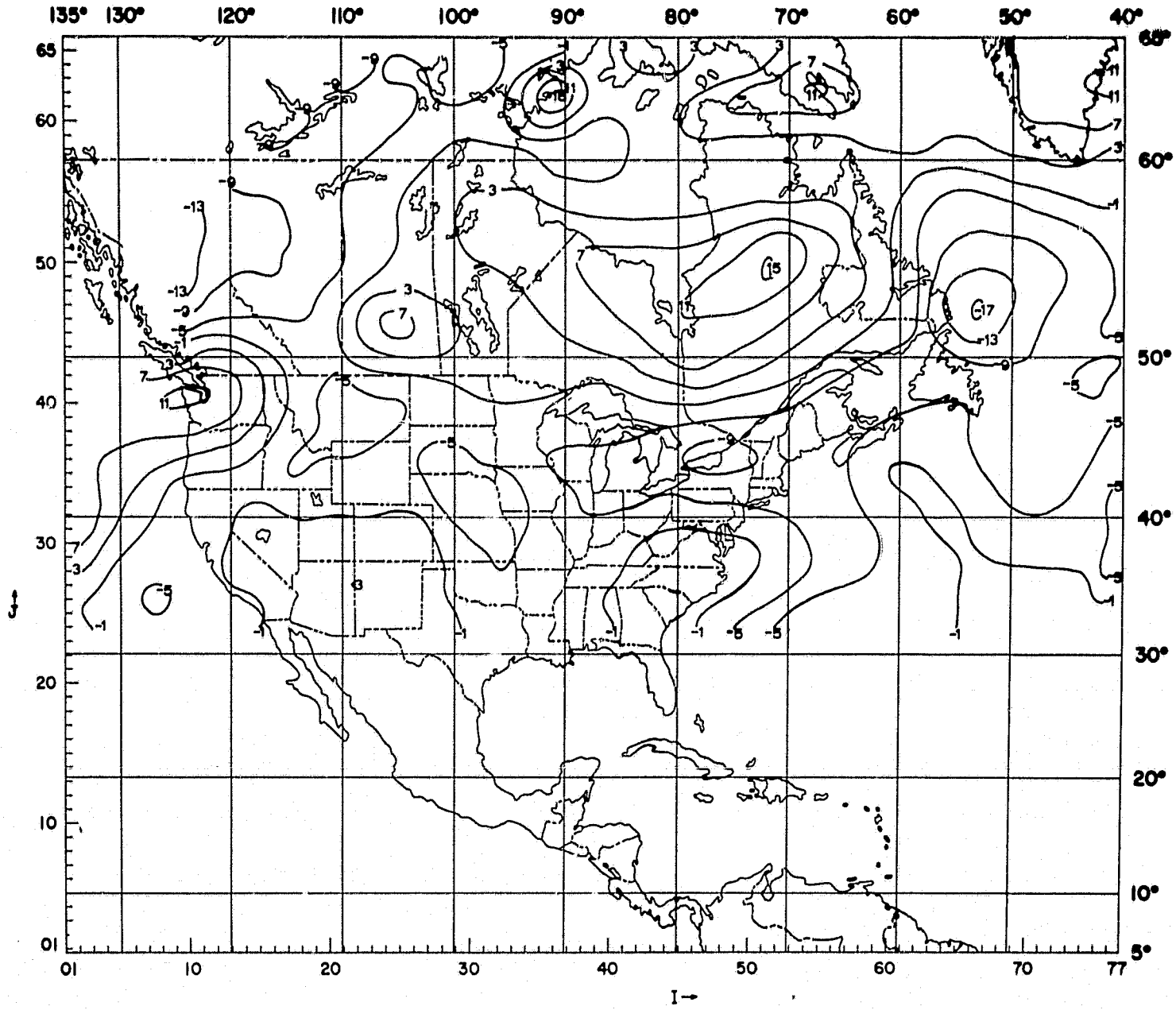

Figure 11–The error field $\epsilon(I, J)$ in decameters for the composite area B–C corresponding to day 4 (00GMT, 19 July 1966). The root mean square error using the optimum four-predictor equation averaged 57.04 gpm over the composite area B–C.